Information Sciences 677 (2024) 120900

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

# Attribute reduction for hierarchical classification based on improved fuzzy rough set

Jie Yang [a,b], Xiaodan Qin [a], Guoyin Wang [a], Qinghua Zhang [a], Shuai Li [d], Di Wu [a,c,*]

[a] *Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China*
[b] *School of Physics and Electronic Science, Zunyi Normal University, Zunyi, 563002, China*
[c] *College of Computer and Information Science, Southwest University, Chongqing, 400715, China*
[d] *School of Mathematics and Information Sciences, Nanchang Hangkong University, Nanchang, 330063, China*

ARTICLE INFO

*Keywords:*
Fuzzy rough set
Stablility factor
Tolerance index
Hierarchical classification
Attribute reduction

ABSTRACT

Attribute reduction plays a critical role in extracting valuable information from high-dimensional datasets. Compared to Pawlak rough set, fuzzy rough set can preserve more data information, making it a prominent focus in the research of attribute reduction. However, current fuzzy rough set-based attribute reduction focuses on flat classification, neglecting distinguishable stability information in hierarchical classification, which leads to insufficient data utilization and reduces the accuracy of attribute reduction. To address these issues, this paper presents two types of fuzzy rough set, named IFRS-I and IFRS-II. Especially, IFRS-I is an improved fuzzy rough set for flat classification, while IFRS-II is constructed based on IFRS-I for hierarchical classification. Unlike traditional fuzzy rough set, IFRS-II has the following two advantages: (1) a stability factor is designed to measure the stability difference among decision classes, (2) a tolerance index is designed to represent the tolerance distance between decision classes based on the approximate information of different decision classes in hierarchical classification. Finally, a stable and effective attribute reduction based on IFRS-II (ARIFRS-II) is designed for hierarchical classification. Experiments demonstrate that compared with the existing related algorithms, ARIFRS-II obtains a higher classification accuracy and stability, while maintaining a suitable number of subsets after reduction.

## 1. Introduction

Rough set [1], as an effective tool for dealing with uncertain concept, has a relatively mature mathematical foundation. However, when the data in a classification problem is continuous, the rough set needs to discretize the data, and it is easy to lose numerous useful information. Fuzzy rough set [2], a special tool to deal with uncertainty problems, is able to process continuous data and has been extensively studied [3–10]. Zhang [3] extended the fuzzy rough set model by introducing the fuzzy coimplication operator and triangular conorm. Ye [4] designed a fuzzy rough set model that satisfies reflexivity using fuzzy logic operators. Zhan [5] proposed a multi-granulation fuzzy rough set model and applied it to multi-criteria group decision-making problems. In addition, fuzzy rough

* Corresponding author at: Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China.

*E-mail addresses:* yj530966074@foxmail.com (J. Yang), S210201082@stu.cqupt.edu.cn (X. Qin), wanggy@cqupt.edu.cn (G. Wang), zhangqh@cqupt.edu.cn (Q. Zhang), zorro0311@126.com (S. Li), wudi.cigit@gmail.com (D. Wu).

https://doi.org/10.1016/j.ins.2024.120900
Received 19 January 2024; Received in revised form 8 March 2024; Accepted 3 June 2024

Available online 7 June 2024

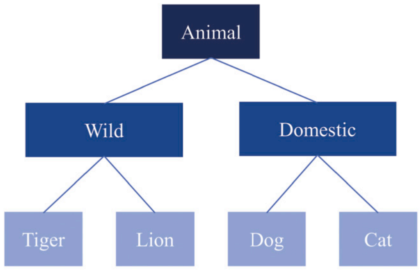



**Fig. 1.** The hierarchical structure of animal classification systems.

set is applied to system monitoring [6,7], pattern recognition [8,9], fuzzy control [10], etc. Due to the use of fuzzy binary relations to represent the similarity between samples in fuzzy rough set, numerical attribute values no longer need to be discretized. This paper chooses fuzzy rough set to the increased utilization of data.

With the maturity of information technology, the providing dataset also includes a lot of extraneous information [11,12]. However, too much redundant information may affect the final evaluation. For example, in the classification problem, redundant information increases the likelihood of incorrect classification results. Therefore, how to eliminate these redundant attributes has become the focus of current research. Attribute reduction aims to simplify the data knowledge base by removing irrelevant or unimportant attributes, simplifying the system without changing the classification ability of the original knowledge system, and facilitating the analysis and processing of data in fuzzy rough set. There is numerous research on attribute reduction based on fuzzy rough set [13–20]. Yuan [13] designed an unsupervised mixed attribute reduction model using fuzzy rough set. Wang [14] measured both the lower and upper approximations of fuzzy decision-making and presented appropriate attribute reduction algorithm. Lin [15] proposed an effective attribute reduction with multi-label learning. Nevertheless, in practical applications, the data itself often contains less emphasized information, namely, the stability varies among different decision classes. As a result, the utilization of information only depends on individual values, overlooking the information brought by the overall characteristics of decision classes. In addressing this issue, we propose a fuzzy rough set by considering stability factor. Furthermore, these works failed to comprehensively account for the inherent information of decision classes, i.e., hierarchical structure. Traditional attribute reduction algorithms assume that each decision class is independent of each other. For example, as shown in Fig. 1, the cost of identifying 'Dog' as 'Lion' is the same as that of classifying 'Dog' as 'Cat' in flat classification. Obviously, the latter classification is generally more acceptable when considering the hierarchical structure, because dogs and cats are both domestic animals.

Granular computing (GrC) [21–24], simulating human cognitive mechanisms, provides a powerful tool for addressing complex problem-solving and managing uncertain information by emphasizing a multifaceted understanding and description of the real world across multiple levels and perspectives. "It has already been applied to many aspects, such as granular mappings [21], three-way decision [22], multi-granularity joint problem-solving mechanisms [24], etc. In the face of super-large number and super-high-dimensional learning problems, from the perspective of GrC, we usually organize them into a hierarchical structure according to the classification way from coarser to finer. Many studies have shown that there is a hierarchical structure among decision classes [25,26]. The use of hierarchical classification provides a lot of effective information, and currently hierarchical classification has been applied to various kinds of research [27–31]. Li [27] applied hierarchical classification to rough set theory based on information entropy and obtained remarkable experimental results. Wang [28] proposed an uncertainty measure for hierarchical classification and further designed a multi-granularity decision model. Qiu [29] proposed a feature selection algorithm that divided feature selection tasks into multiple subtasks through a hierarchical structure. Zhao [30] proposed a feature selection algorithm based on sibling nodes through hierarchical classification. Moreover, hierarchical classification has also been used for protein function prediction [32], image classification [33], text categorization [34], etc. In reality, there are certain dependencies among different decision classes. When the distance among decision classes is small, they become more challenging to distinguish. In such cases, if two decision classes have a high approximation degree, the cost of grouping them together is lower, and vice versa. Therefore, the influence of approximation degree is negatively correlated with the distance between decision classes. However, this aspect has not been thoroughly explored in current research, leading to unnecessary computations. For instance, when the differences between decision classes are significant, it may not be practical to classify them into the same category, but similar calculation rules are still applied. This oversight neglects the dynamic connection between approximation degree and decision class differences. In this paper, we employ the hierarchical quotient space structure (HQSS) to characterize the hierarchical classification, then design the tolerance index to represent the tolerance distance between decision classes, and further design a novel fuzzy rough set based on IFRS-I, named IFRS-II. Furthermore, we design an attribute reduction algorithm based on IFRS-II, named ARIFRS-II.

The contributions of this paper are as follows: (1) According to the stability difference between decision classes, the stability factor is formulated to design IFRS-I. (2) Considering the approximation degree of decision class in hierarchical structure, the tolerance index is devised to construct IFRS-II based on IFRS-I. (3) An attribute reduction for hierarchical classification based on IFRS-II is designed.

The arrangement of this paper is as follows: Section 2 briefly introduces relative core concepts. Section 3 introduces a stability factor into fuzzy rough set model to construct IFRS-I. Section 4 introduces hierarchical classification to IFRS-I to further construct IFRS-II by considering tolerance index. The corresponding experiments are carried out in Section 5. Section 6 provides the conclusion.

## 2. Preliminaries

In this section, some basic concepts related to fuzzy rough set are presented to simplify the structure of the full paper, and Definition 6 is presented to calculate the distance between two decision classes. Let $S = (U, AT \cup D)$ be a decision information system, where $U = \{x_1, x_2, ..., x_N\}$ is called the universe of discourse, $AT$ and $D$ represent conditional attributes and decision attribute, respectively.

**Definition 1.** [35] Let $S = (U, AT \cup D)$ be a decision information system and $R_B$ be a similarity relation on $U$ under conditional attribute $B$, $B \subseteq AT$. $R$ is an equivalence relation induced by decision attribute $D$ and $U$ is divided into $m$ crisp decision classes by $R$, expressed as $U/R = \{Q_1, Q_2, ..., Q_m\}$. The lower approximation $\underline{R}_B$ and upper approximation $\overline{R}_B$ of $Q_i$ in fuzzy rough set are defined as follows:

$$\underline{R}_B(Q_i)(x) = \min_{y \notin Q_i} \left(1 - R_B(x,\ y)\right), \quad x \in U \tag{1}$$

$$\overline{R}_B(Q_i)(x) = \max_{y \in Q_i} R_B(x,\ y), \quad x \in U \tag{2}$$

**Definition 2.** [27] Let $S = (U, AT \cup D)$ be a decision information system, $L = \{l_1, l_2, ..., l_z\}$ be a label set and $v : L \to 2^U$ be an injective mapping. For $\forall l_i, l_j \in L$, if $v(l_i) \subseteq v(l_j)$, we can define the partial order relation $l_i \preceq l_j$. If the label set $L$ meets the following conditions:
(1) $\exists l_0 \in L$ satisfying $U = v(l_0)$;
(2) $v(l_i) \cap v(l_j) \neq \emptyset \Rightarrow l_i \preceq l_j \ or\ l_j \preceq l_i, \quad \forall l_i,\ l_j \in L$;
(3) For the leaf label set $L' = \{l \in L | l' \nprec l, \forall l' \in L\}$, $\bigcup_{l \in L'} v(l) = U$.
Then, the triple $H = \langle U, L, v\rangle$ is a hierarchical classification.

**Definition 3.** [27] Let $H = \langle U, L, v\rangle$ be a hierarchical classification and $T(L, \preceq)$ be a corresponding tree structure, where $L = \{l_1, l_2, ..., l_z\}$ be a label set and $\preceq$ be the partial order relation. For each leaf label not in the deepest layer, it is copied into each subsequent layer as its own child label until the deepest layer is reached to form a new tree structure $T'(L, \preceq)$. All labels are replaced with the corresponding subset of samples according to decision value of $H$ to form the hierarchical quotient space structure $HQ(U, \mathfrak{R})$. Then $HQ(U, \mathfrak{R})$ is called the hierarchical quotient space structure (HQSS) representation of hierarchical classification $H$.

**Definition 4.** [36,37] Let $S = (U, AT \cup D)$ be a decision information system, $\mathfrak{R} = \{R_h | h = 0, 1, ..., z\}$ be a series of equivalence relations on $U$ satisfying $R_h \prec R_{h-1}, h = 1, ..., z$. Then, the collection of quotient sets $HQ(U, \mathfrak{R}) = \{U/R_h | h = 0, 1, ..., z\}$ is called the hierarchical quotient space structure ($HQSS$) of $\mathfrak{R}$ on $U$. $U$ is divided into $n$ crisp decision classes and $m$ crisp decision classes by $R_1$ and $R_2$, expressed as $U/R_1 = \{P_1, P_2, ..., P_n\}$ and $U/R_2 = \{Q_1, Q_2, ..., Q_m\}$. The knowledge distance between $U/R_1$ and $U/R_2$ is defined as follows:

$$KD(U/R_1, U/R_2) = \frac{1}{|U|} \sum_{i=1}^{n} \sum_{j=1}^{m} d_{ij} f_{ij} \tag{3}$$

where $d_{ij} = \frac{\left|P_i \cup Q_j\right| - \left|P_i \cap Q_j\right|}{|U|}$ and $f_{ij} = \left|P_i \cap Q_j\right|$.

**Definition 5.** [38] Let $S = (U, AT \cup D)$ be a decision information system. $X$, $Y$ and $Z$ are three finite sets on $U$. $\Delta(\cdot, \cdot)$ is a distance measure if it satisfies the following conditions:
(1) Positive: $\Delta(X, Y) \geq 0$;
(2) Symmetric: $\Delta(X, Y) = \Delta(Y, X)$;
(3) Triangle inequality: $\Delta(X, Y) + \Delta(Y, Z) \geq \Delta(X, Z)$.

As a common distance measure, the Euclidean distance is frequently utilized in fuzzy rough set to determine the difference between two samples. The formula is as follows:

$$d(x_a, x_b) = \sqrt{\sum_{k=1}^{t} (x_{ak} - x_{bk})^2} \tag{4}$$

where $x_a = \left\{x_{a1},\ x_{a2}, ...,\ x_{at}\right\}$ and $x_b = \left\{x_{b1},\ x_{b2}, ...,\ x_{bt}\right\}$ are two samples in Euclidean $n$-space, and $x_{ik}$ represents the value under the $k-th$ attribute of sample $x_i$.

In this paper, Euclidean distance is introduced as the basis for calculating the distance of decision classes. To avoid losses caused by the protrusion of individual samples, the average Euclidean distance between two decision classes is defined as the distance between the two decision classes.

**Definition 6.** Let $Q_1$ and $Q_2$ represent two decision classes, then the distance between $Q_1$ and $Q_2$ under conditional attribute $B$ is defined as follows:

$$Dis_B(Q_1,Q_2)=\frac{\frac{1}{|Q_1|}\sum_{i=1}^{|Q_1|}d(x_i,\overline{y})+\frac{1}{|Q_2|}\sum_{j=1}^{|Q_2|}d(\overline{x},y_j)}{2} \tag{5}$$

where $x_i\in Q_1$, $y_j\in Q_2$, $\overline{x}$ and $\overline{y}$ are the mean points of $Q_1$ and $Q_2$, $|Q_1|$ and $|Q_2|$ represent the number of samples in decision classes $Q_1$ and $Q_2$.

**Theorem 1.** *$Dis_B(\cdot,\cdot)$ is a distance measure.*

**Proof.** Suppose $Q_1$, $Q_2$ and $Q_3$ are three decision classes, $x_i\in Q_1$, $y_i\in Q_2$, $z_i\in Q_3$, $\overline{x}$, $\overline{y}$ and $\overline{z}$ are the mean points of $Q_1$, $Q_2$ and $Q_3$, respectively.
(1) Positive and symmetric:
It is obvious that $Dis_B(Q_1,Q_2)\geq 0$ and $Dis_B(Q_1,Q_2)=Dis_B(Q_2,Q_1)$
(2) Triangle inequality:

$$Dis_B(Q_1,Q_2)+Dis_B(Q_2,Q_3)=\frac{\frac{1}{|Q_1|}\sum_{i=1}^{|Q_1|}d(x_i,\overline{y})+\frac{1}{|Q_2|}\sum_{i=1}^{|Q_2|}d(\overline{x},y_i)+\frac{1}{|Q_2|}\sum_{i=1}^{|Q_2|}d(y_i,\overline{z})+\frac{1}{|Q_3|}\sum_{i=1}^{|Q_3|}d(\overline{y},z_i)}{2},$$

$$Dis_B(Q_1,Q_3)=\frac{\frac{1}{|Q_1|}\sum_{i=1}^{|Q_1|}d(x_i,\overline{z})+\frac{1}{|Q_3|}\sum_{i=1}^{|Q_3|}d(\overline{x},z_i)}{2}.$$

According to Cauchy's inequality:

$$\sum_{i=1}^{n}a_i^2\times\sum_{i=1}^{n}b_i^2\geq(\sum_{i=1}^{n}a_ib_i)^2,$$

$$(\sqrt{\sum_{i=1}^{n}a_i^2})^2+(\sqrt{\sum_{i=1}^{n}b_i^2})^2+2\sqrt{\sum_{i=1}^{n}a_i^2\times\sum_{i=1}^{n}b_i^2}\geq\sum_{i=1}^{n}a_i^2+\sum_{i=1}^{n}b_i^2+2\sum_{i=1}^{n}a_ib_i,$$

$$\sqrt{\sum_{i=1}^{n}a_i^2}+\sqrt{\sum_{i=1}^{n}b_i^2}\geq\sqrt{\sum_{i=1}^{n}(a_i+b_i)^2}.$$

Easy to get: $\sum_{i=1}^{m}\sqrt{\sum_{j=1}^{n}x_{ij}^2}\geq\sqrt{\sum_{j=1}^{n}(\sum_{i=1}^{m}x_{ij})^2}$. 1)
According to 1):

$$\frac{1}{|Q_2|}\sum_{i=1}^{|Q_2|}d(y_i,\overline{z})=\frac{1}{|Q_2|}\sum_{i=1}^{|Q_2|}\sqrt{\sum_{j=1}^{n}(y_{ij}-\overline{z}_j)^2}\geq\frac{1}{|Q_2|}\sqrt{\sum_{j=1}^{n}(\sum_{i=1}^{|Q_2|}y_{ij}-|Q_2|\times\overline{z}_j)^2}=\sqrt{\sum_{j=1}^{n}(\overline{y}_j-\overline{z}_j)^2}=d(\overline{y},\overline{z}),$$

where $y_{ij}$ represents the value under the $j-th$ attribute of sample $y_i$, $\overline{y}_j$ and $\overline{z}_j$ represent the value under the $j-th$ attribute of $\overline{y}$ and $\overline{z}$.

We have: $\frac{1}{|Q_1|}\sum_{i=1}^{|Q_1|}d(x_i,\overline{y})+\frac{1}{|Q_2|}\sum_{i=1}^{|Q_2|}d(y_i,\overline{z})\geq\frac{1}{|Q_1|}\sum_{i=1}^{|Q_1|}d(x_i,\overline{y})+d(\overline{y},\overline{z})$. 2)
Because $d(x,y)$ is the Euclidean distance, which satisfies the triangle inequality.

$$\frac{1}{|Q_1|}\sum_{i=1}^{|Q_1|}d(x_i,\overline{y})+d(\overline{y},\overline{z})=\frac{1}{|Q_1|}(\sum_{i=1}^{|Q_1|}(d(x_i,\overline{y})+d(\overline{y},\overline{z})))\geq\frac{1}{|Q_1|}\sum_{i=1}^{|Q_1|}(d(x_i,\overline{z})),\ 3)$$

Combining 2) and 3), we have: $\frac{1}{|Q_1|}\sum_{i=1}^{|Q_1|}d(x_i,\overline{y})+\frac{1}{|Q_2|}\sum_{i=1}^{|Q_2|}d(y_i,\overline{z})\geq\frac{1}{|Q_1|}\sum_{i=1}^{|Q_1|}(d(x_i,\overline{z}))$.

Similarly, $\frac{1}{|Q_3|}\sum_{i=1}^{|Q_3|}d(z_i,\overline{y})+\frac{1}{|Q_2|}\sum_{i=1}^{|Q_2|}d(y_i,\overline{x})\geq\frac{1}{|Q_3|}\sum_{i=1}^{|Q_3|}(d(z_i,\overline{x}))$.
Then, $Dis_B(Q_1,Q_2)+Dis_B(Q_2,Q_3)\geq Dis_B(Q_1,Q_3)$.
Therefore, $Dis_B(\cdot,\cdot)$ is a distance measure. □

## 3. Construction of fuzzy rough set by considering stability factor

As well know, the most fuzzy rough set models are established based on independent sample, which is quite dependent on the quality of the sample itself, and the calculation is more complex. In addition, in traditional fuzzy rough set, the lower approximation can be understood as the minimum difference between heterogeneous samples. This paper constructs the fuzzy rough set model based on sample set to increase the anti-interference ability of the model. We represent the sample set in terms of the decision class and use the minimum difference between heterogeneous decision classes to represent the lower approximation. This section presents some definitions and examples before and after the introduction of stability factor.

**Table 1**
Decision information system of subject classification.

| sample | score | decision |
|---|---|---|
| $x_1$ | 0.75 | Chinese |
| $x_2$ | 0.83 | Chinese |
| $x_3$ | 0.92 | Chinese |
| $x_4$ | 0.58 | History |
| $x_5$ | 0.63 | History |
| $x_6$ | 0.75 | History |
| $x_7$ | 0.00 | Mathematics |
| $x_8$ | 0.75 | Mathematics |
| $x_9$ | 1.00 | Mathematics |

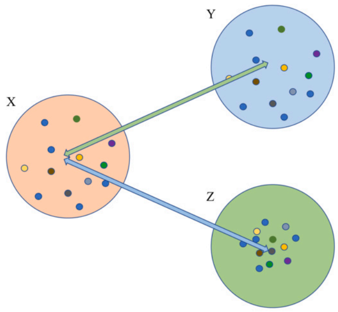



**Fig. 2.** Comparison of difference among three sample sets.

**Definition 7.** Let $S = (U, AT \cup D)$ be a decision information system, where $U$ is divided into $m$ crisp decision classes by the equivalence relation $R$ induced by $D$, expressed as $U/R = \{Q_1, Q_2, ..., Q_m\}$, $B \subseteq AT$. The lower approximation $\underline{R_B}$ of $Q_i$ under conditional attribute $B$ in fuzzy rough set is defined as follows:

$$\underline{R}_B(Q_i) = \min_{i \neq j} \left(1 - e^{-Dis_B(Q_i, Q_j)}\right), \quad Q_i, Q_j \in U/R \tag{6}$$

In practical application, samples with different labels usually have different stability, but few studies have paid attention to this point. For example, in daily life, there is not a subtle difference in Chinese language during exams, but there is a significant difference in Mathematics. Here is an example to illustrate the current disadvantage without considering stability differences.

**Example 1.** Suppose that in the decision information system of subject classification, one of the conditional attributes is score, whose values are 85, 90, 95, 75, 78, 85, 40, 85, 100, and there are three decision classes, which are Chinese, History and Mathematics subjects, respectively. The normalized decision information table is shown in Table 1. The distance from sample $x_2$ to sample $x_6$ and from sample $x_2$ to sample $x_8$ can be obtained by using formula (4). The calculation process is as follows:

$$d(x_2, x_6) = \sqrt{(0.83 - 0.75)^2} = 0.08,$$

$$d(x_2, x_8) = \sqrt{(0.83 - 0.75)^2} = 0.08.$$

In fact, it is easy to observe that the score distribution of Chinese and Historical is more dense, while the score distribution of Mathematical is scattered. However, the $x_6$ and $x_8$ values are consistent, resulting in the consistent distances between $x_2$ to $x_6$ and $x_2$ to $x_8$ calculated based on Euclidean distance. This loses a key information inherent in the decision information system, which is the stability of the data distribution. As shown in Fig. 2, $X$, $Y$ and $Z$ denote the three sample sets, respectively. If the dispersion degree of the sample set is not considered, the distance from $X$ to $Y$ is the same as the distance from $X$ to $Z$, which ignores the stability difference of the sample set itself. Actually, the difference between $X$ and $Z$ is greater than the difference between $X$ and $Y$. This is because $X$ and $Y$ have a more dispersed data distribution while $Z$ has a denser data distribution. Therefore, it fails to reflect this difference between sample sets effectively by only calculating their Euclidean distance. In order to increase the utilization of our data, it is necessary to have a factor that characterizes the stability of the sample set. Standard deviation, the most commonly used statistic in statistics, reflects the dispersion of a dataset. This paper utilizes this property to construct a stability factor, which is defined as follows:

**Definition 8.** Assume that $\left|\sigma_{Q_1} - \sigma_{Q_2}\right|$ stands for absolute value of $\sigma_{Q_1} - \sigma_{Q_2}$, $\sigma_{Q_1}$ and $\sigma_{Q_2}$ represents the standard deviation of the decision classes $Q_1$ and $Q_2$, then the stability factor $\mu_{12}$ between decision classes $Q_1$ and $Q_2$ is defined as follows:

$$\mu_{12} = 1 + \frac{\left|\sigma_{Q_1} - \sigma_{Q_2}\right|}{\max(\sigma_{Q_1}, \sigma_{Q_2})} \tag{7}$$

The following formula is obtained by combining the stability factor with the Euclidean distance:

$$d_\mu(x_a, x_b) = \sqrt{\sum_{k=1}^{t} (x_{ak} - x_{bk})^2 \times \mu_{12}} \tag{8}$$

where $x_a$ belongs to the decision class $Q_1$ and $x_b$ belongs to the decision class $Q_2$.

**Example 2.** (Continued Example 1) Suppose that Chinese, History, and Mathematics are represented by $Q_1$, $Q_2$, and $Q_3$, respectively. After introducing the stability factor, the distance from $x_2$ to $x_6$ and from $x_2$ to $x_8$ are calculated as follows:

$$\begin{aligned}
\mu_{12} &= 1 + \frac{\left|\sigma_{Q_1} - \sigma_{Q_2}\right|}{\max(\sigma_{Q_1}, \sigma_{Q_2})} = 1 + \frac{|0.085 - 0.087|}{0.087} \approx 1.02,\\
\mu_{13} &= 1 + \frac{\left|\sigma_{Q_1} - \sigma_{Q_3}\right|}{\max(\sigma_{Q_1}, \sigma_{Q_3})} = 1 + \frac{|0.085 - 0.520|}{0.520} \approx 1.84,\\
d_\mu(x_2, x_6) &= \sqrt{(0.83 - 0.75)^2 \times \mu_{12}} \approx 0.08,\\
d_\mu(x_2, x_8) &= \sqrt{(0.83 - 0.75)^2 \times \mu_{13}} \approx 0.11.
\end{aligned}$$

The introduction of stability factor $\mu$ makes $d_\mu(x_2, x_6) < d_\mu(x_2, x_8)$. Comparing with Example 1, the introduction of stability factor makes the calculation of differences between different samples more realistic and improves the performance of classification. Therefore, we design a fuzzy rough set considering the stability factor (IFRS-I), and the lower approximation based on the decision class in Definition 7 can be redefined as follows:

**Definition 9.** Let $S = (U, AT \cup D)$ be a decision information system, where $U$ is divided into $m$ crisp decision classes by the equivalence relation $R$ induced by $D$, expressed as $U/R = \{Q_1, Q_2, ..., Q_m\}$, $B \subseteq AT$. The lower approximation $\underline{R}_B^\mu$ of $Q_i$ under conditional attribute $B$ in IFRS-I is defined as follows:

$$\underline{R}_B^\mu(Q_i) = \min_{i \neq j} \left(1 - e^{-Dis_B^\mu(Q_i, Q_j)}\right), \quad Q_i, Q_j \in U/R \tag{9}$$

where $Dis_B^\mu(Q_1, Q_2) = \frac{\frac{1}{|Q_1|}\sum_{i=1}^{|Q_1|} d_\mu(x_i, \overline{y}) + \frac{1}{|Q_2|}\sum_{i=1}^{|Q_2|} d_\mu(\overline{x}, y_i)}{2}$.

Based on Definition 9, we can calculate the dependence of decision attribute $D$ on the conditional attribute subset $B$ under IFRS-I, defined as follows:

**Definition 10.** Let $S = (U, AT \cup D)$ be a decision information system, where $U$ is divided into $m$ crisp decision classes by the equivalence relation $R$ induced by $D$, expressed as $U/R = \{Q_1, Q_2, ..., Q_m\}$, $B \subseteq AT$, $\underline{R}_B^\mu(Q_i)$ is the lower approximation of $Q_i$ in IFRS-I, then the dependence $\gamma_B$ of the decision attribute under conditional attribute $B$ is defined as follows:

$$\gamma_B = \frac{\sum_{i=1}^{m} \underline{R}_B^\mu(Q_i)}{m} \tag{10}$$

Here is an example to show the calculation of $\gamma_B$:

**Example 3.** (Continued Example 2) The condition score is represented by $B$, and the dependence of decision attribute $D$ under conditional attribute $B$ is calculated as follows:

$$\begin{aligned}
&Dis_B^\mu(Q_1, Q_2) = 0.18,\ Dis_B^\mu(Q_1, Q_3) = 0.41,\ Dis_B^\mu(Q_2, Q_3) = 0.30,\\
&\underline{R}_B^\mu(Q_1) = \min\left(1 - e^{-Dis_B^\mu(Q_1, Q_2)}, 1 - e^{-Dis_B^\mu(Q_1, Q_3)}\right) = 0.16,\\
&\underline{R}_B^\mu(Q_2) = \min\left(1 - e^{-Dis_B^\mu(Q_2, Q_1)}, 1 - e^{-Dis_B^\mu(Q_2, Q_3)}\right) = 0.16,\\
&\underline{R}_B^\mu(Q_3) = \min\left(1 - e^{-Dis_B^\mu(Q_3, Q_1)}, 1 - e^{-Dis_B^\mu(Q_3, Q_2)}\right) = 0.26,
\end{aligned}$$

**Table 2**
Decision information system for animal classification after normalization.

| sample | Food intake ($c_1$) | Tail length ($c_2$) | Hair length ($c_3$) | Running speed ($c_4$) | Running speed' ($c_4{}'$) | Life span ($c_5$) | Height ($c_6$) | decision |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | 0.02 | 0.00 | 0.10 | 0.00 | 0.00 | 0.50 | 0.09 | Dog |
| $x_2$ | 0.04 | 0.01 | 0.10 | 0.67 | 0.67 | 0.50 | 0.13 | Dog |
| $x_3$ | 0.04 | 0.08 | 0.30 | 0.25 | 0.25 | 0.00 | 0.16 | Dog |
| $x_4$ | 0.75 | 1.00 | 0.40 | 0.57 | 0.57 | 0.50 | 0.74 | Tiger |
| $x_5$ | 0.62 | 0.76 | 0.50 | 0.53 | 0.53 | 0.25 | 0.79 | Tiger |
| $x_6$ | 0.87 | 0.88 | 0.50 | 0.63 | 0.63 | 0.88 | 0.70 | Tiger |
| $x_7$ | 1.00 | 0.65 | 1.00 | 0.83 | 0.20 | 1.00 | 0.83 | Lion |
| $x_8$ | 0.94 | 1.00 | 0.80 | 1.00 | 0.42 | 0.75 | 0.96 | Lion |
| $x_9$ | 0.81 | 0.76 | 0.70 | 0.97 | 0.33 | 0.63 | 1.00 | Lion |
| $x_{10}$ | 0.00 | 0.06 | 0.00 | 0.20 | 0.83 | 0.13 | 0.01 | Cat |
| $x_{11}$ | 0.00 | 0.04 | 0.30 | 0.42 | 1.00 | 0.50 | 0.00 | Cat |
| $x_{12}$ | 0.00 | 0.08 | 0.40 | 0.33 | 0.97 | 0.75 | 0.02 | Cat |

We have $\gamma_B = \frac{\sum_{i=1}^{3} \underline{R}^{\mu}_B(Q_i)}{3} = 0.19$. In order to compare the difference before and after the introduction of stability factor, we replace $Dis^{\mu}_B$ in the $\gamma_B$ formula with $Dis_B$, expressed by $\gamma'_B$, and calculate as follows:

$$Dis_B(Q_1,Q_2) = 0.18,\ Dis_B(Q_1,Q_3) = 0.31,\ Dis_B(Q_2,Q_3) = 0.22,$$
$$\underline{R}_B(Q_1) = 0.16,\ \underline{R}_B(Q_2) = 0.16,\ \underline{R}_B(Q_3) = 0.20,$$
$$\gamma'_B = 0.17 < \gamma_B.$$

It can be seen that due to the different stability of decision classes in $Q_1$, $Q_2$, and $Q_3$, the distance from $Q_1$ to $Q_3$ and $Q_1$ to $Q_2$ increases after introducing stability factor, leading to an increase in dependency. This indicates that we have successfully utilized stability information other than numerical values in our data.

## 4. Attribute reduction based on IFRS-II for hierarchical classification

Traditional flat classification assumes that all decision classes are independent of each other, resulting in the loss of the dependencies of different decision classes in decision information system. Hierarchical classification solves this problem by considering the hierarchical structure of decision classes and linking the information of decision classes. Based on the fuzzy rough set model with stability factor introduced in the previous section, this section defines the fuzzy rough set model with hierarchical classification introduced. The following example demonstrates the disadvantages of flat classification.

**Example 4.** As is shown in Table 2, $AT = \{c_1, c_2, c_3, c_4, c_4{}', c_5, c_6\}$, the last column of the table is decision attribute $D$. We use symbols to represent these four decision classes, that is, $Q_1 \Leftrightarrow \{'Dog'\}$, $Q_2 \Leftrightarrow \{'Tiger'\}$, $Q_3 \Leftrightarrow \{'Lion'\}$ and $Q_4 \Leftrightarrow \{'Cat'\}$. Swap the attribute values of 'Lion' and 'Cat' in the conditional attribute $c_4$, and in order to distinguish the changed conditional attribute, we represent it as $c_4{}'$. We compare the dependence of decision attribute $D$ in two cases before and after conditional attribute change.

$$\underline{R}^{\mu}_{c_4}(Q_1) = 0.19,\quad \underline{R}^{\mu}_{c_4}(Q_2) = 0.28,\quad \underline{R}^{\mu}_{c_4}(Q_3) = 0.35,\quad \underline{R}^{\mu}_{c_4}(Q_4) = 0.19,$$
$$\underline{R}^{\mu}_{c_4{}'}(Q_1) = 0.19,\quad \underline{R}^{\mu}_{c_4{}'}(Q_2) = 0.28,\quad \underline{R}^{\mu}_{c_4{}'}(Q_3) = 0.19,\quad \underline{R}^{\mu}_{c_4{}'}(Q_4) = 0.35.$$

We have $\gamma_{c_4} = \gamma_{c_4{}'} = 0.25$.

According to Fig. 1, it is easy to find that the four decision classes in the decision system of Table 2 are related. 'Dog' and 'Cat' are domestic animals, and 'Lion' and 'Tiger' are wild animals. However, Example 4 obtains the same dependence of decision attribute $D$ after exchanging the attribute values of the samples labeled as 'Lion' and 'Cat' under conditional attribute $c_4$, which also proves that flat classification fails to distinguish the dependency relationship to describe the decision class itself. The fuzzy rough set considering stability factor (IFRS-I) proposed in this paper enhances the utilization of original data information by introducing a stability factor. However, IFRS-I considers the flat structure. In order to enable the widespread application of IFRS-I in various scenarios, such as the case where there is a connection between decision classes, this paper introduces a hierarchical classification based on IFRS-I, resulting in improved fuzzy rough set (IFRS-II). To better describe the hierarchical classification, throughout the rest of this paper, we use the hierarchical quotient space structure (HQSS) to represent hierarchical classification of decision classes. The following shows the process of the HQSS representation of hierarchical classification. Then, Example 6 illustrates the change in dependency after the introduction of hierarchical classification.

**Example 5.** (Continued Example 4) As shown in Fig. 3, we used the decision information system in Table 2 and set $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}$ as the dataset. To illustrate the conversion process when the original leaf label is not in the same layer, we remove the 'Cat' and 'Dog' label. The label set of tree structure $T(L, \preceq)$ is $L = \{'Animal', 'Wild', 'Domestic',$

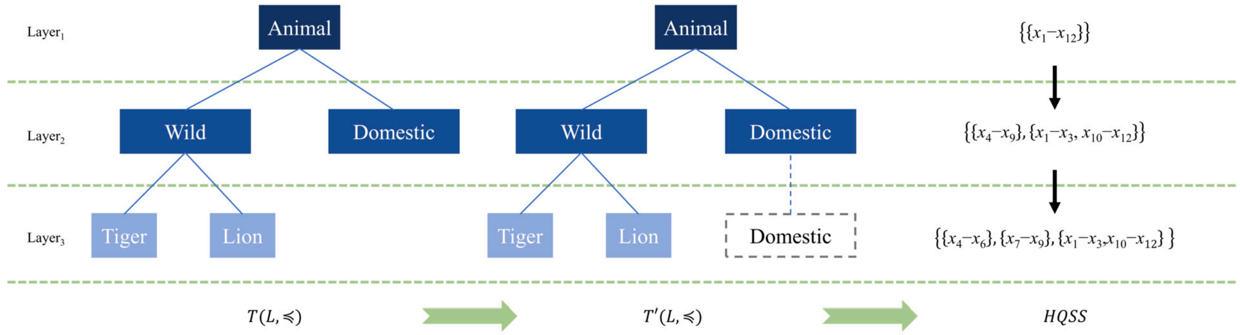



**Fig. 3.** HQSS representation of hierarchical classification.

$'Tiger','Lion'\}$. The leaf labels and corresponding sample are $\{x_1-x_3, x_{10}-x_{12}\}-'Domestic'$, $\{x_4-x_6\}-'Tiger'$, $\{x_7-x_9\}-'Lion'$. There are three layers, the leaf label 'Domestic' is not at the deepest layer and copies itself as a sub-label until all the leaf labels are at the deepest level to form tree structure $T'(L,\preceq)$. Finally, replace labels with corresponding samples to form the HQSS and the process is shown in Fig. 3.

According to Example 4, if the hierarchical classification is not considered to calculate the difference among decision classes, the information about the data structure will be lost. Li [27] designed the average approximation degree for samples according to hierarchical classification. Inspired by this, we designed formulas for calculating the approximation degree $\eta_{ij}$ of different sample sets, where $\eta_{ij}=\frac{KD(Lay(LCA(l_{Q_i}^T,l_{Q_j}^T)),\,Lay(l_R^T))}{KD(Lay(l_{depth}^T),Lay(l_0^T))}$, $l_0$, $l_{Q_i}$, $l_{Q_j}$ and $l_{depth}$ represent the root label, the label where the decision class $Q_i$ resides, the label where the decision class $Q_j$ resides, and label of deepest layer, respectively. $l^T$ stands for placing the label $l$ in the tree structure $T$, $Lay(l^T)$ stands for the layer in $T$ where label $l$ resides, $LCA(l_{Q_i}^T,l_{Q_j}^T)$ stands for the lowest common ancestor of label $l_{Q_i}$ and $l_{Q_j}$ in $T$ and $KD(Lay(.),Lay(.))$ stands for the knowledge distance between two different layers in HQSS representation of tree structure $T$. As $Dis^{\mu}(Q_i,Q_j)$ decreases, it becomes easier to recognize $Q_i$ and $Q_j$ as the same class, making $\eta_{ij}$ more significant, and vice versa. Therefore, we designed the tolerance index to make our fuzzy rough set more suitable for hierarchical classification.

**Definition 11.** Let $S=(U,AT\cup D)$ be a decision information system, where $U$ is divided into $m$ crisp decision classes by the equivalence relation $R$ induced by $D$, expressed as $U/R=\{Q_1,Q_2,...,Q_m\}$, $Q_i,Q_j\in U/R$, $B\subseteq AT$, then the tolerance index of $Q_i$ and $Q_j$ under conditional attribute $B$ in hierarchical classification is defined as follows:

$$\eta_{ij}^H=1+e^{-Dis_B^{\mu}(Q_i,Q_j)}\times\eta_{ij} \tag{11}$$

From Definition 11, it is easy to obtain the following two theorems.

**Theorem 2.** *If $\eta$ is constant, $\eta^H$ decreases with the increase of $Dis^{\mu}$.*

**Proof.** It is easy to observe that when $\eta$ is a constant and $Dis^{\mu}$ is a variable, $\eta^H$ is a decreasing function. Therefore, $\eta^H$ decreases as $Dis^{\mu}$ increases. □

**Theorem 3.** *If $Dis^{\mu}$ is constant, $\eta^H$ increases with the increase of $\eta$.*

**Proof.** It is easy to observe that when $Dis^{\mu}$ is a constant and $\eta$ is a variable, $\eta^H$ is an increasing function. Therefore, $\eta^H$ increases as $\eta$ increases. □

After introducing the tolerance index $\eta_{ij}^H$, IFRS-II is obtained, and the lower approximation in IFRS-II is defined as follows:

**Definition 12.** Let $S=(U,AT\cup D)$ be a decision information system, where $U$ is divided into $m$ crisp decision classes by the equivalence relation $R$ induced by $D$, expressed as $U/R=\{Q_1,Q_2,...,Q_m\}$, $B\subseteq AT$. The lower approximation $\underline{R}_B^{\mu\eta}$ of $Q_i$ under conditional attribute $B$ in IFRS-II is defined as follows:

$$\underline{R}_B^{\mu\eta}(Q_i)=\min_{i\neq j}\left(1-e^{-Dis_B^{\mu\eta}(Q_i,Q_j)}\right),\quad Q_i,Q_j\in U/R \tag{12}$$

where $Dis_B^{\mu\eta}(Q_i,Q_j)=Dis_B^{\mu}(Q_i,Q_j)\times\eta_{ij}^H$.

**Algorithm 1** The calculation of $\gamma_B^{\mu\eta}$.

**Input:** Universe $U$, conditional attribute $B$ and equivalence relation $R$ induced by decision attribute $D$;
**Output:** The dependence $\gamma_B^{\mu\eta}$;

```
for Q_i ∈ U/R do
    for Q_j ∈ U/R, i ≠ j do
        Calculate Dis_B^{μη}(Q_i, Q_j) = η_ij^H × Dis_B^μ(Q_i, Q_j);
    end for
end for
for Q_i ∈ U/R do
    Calculate R_B^{μη}(Q_i) = min_{i≠j} (1 − e^{Dis_B^{μη}(Q_i, Q_j)});
end for
m = size(U/R);
Calculate γ_B^{μη} = (Σ_{i=1}^{m} R_B^{μη}(Q_i)) / m;
```

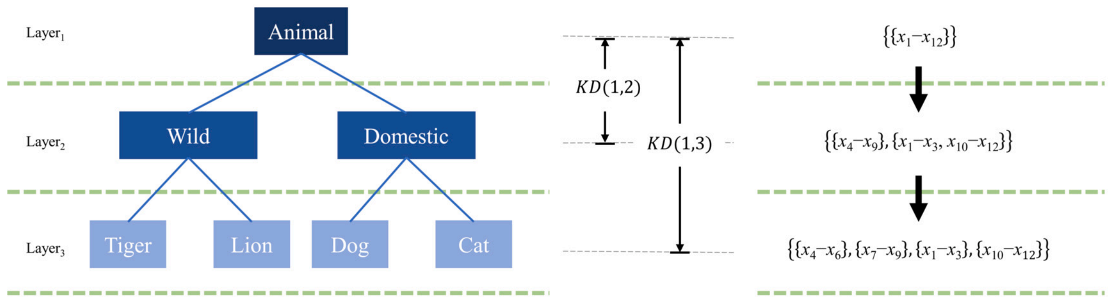



**Fig. 4.** HQSS representation of Fig. 1.

Similarly, based on the approximation formula under IFRS-II, we redefined the dependence of the decision attribute on the conditional attribute $B$ to $\gamma_B^{\mu\eta} = \frac{\sum_{i=1}^{m} \underline{R}_B^{\mu\eta}(Q_i)}{m}$, which is computed by Algorithm 1. The time complexity of the algorithm is $O(n^2)$. Since we calculate based on decision classes, the number of decision classes is much less than the number of samples. Compared with the traditional calculation of dependence based on independent samples, our algorithm is more efficient.

**Example 6.** (Continued Example 4) Fig. 4 shows the HQSS used in the decision system in Table 2. The tolerance indexes between different decision classes can be calculated as follows:

$$\eta_{12} = \eta_{13} = \eta_{24} = \eta_{34} = \frac{KD(1,1)}{KD(1,3)} = 0,$$
$$\eta_{14} = \eta_{23} = \frac{KD(1,2)}{KD(1,3)} = \frac{2}{3},$$
$$\eta_{12}^H = \eta_{13}^H = \eta_{24}^H = \eta_{34}^H = 1,$$
$$\eta_{14}^H = 1 + \frac{2}{3} \times e^{-Dis_c^{\mu}(Q_1,Q_4)},$$
$$\eta_{23}^H = 1 + \frac{2}{3} \times e^{-Dis_c^{\mu}(Q_2,Q_3)}.$$

The aforementioned computation uses the conditional attribute $c$ to represent either $c_4$ or $c_4'$, and after the tolerance index is introduced, the dependence of $c_4$ and $c_4{}'$ are calculated as follows:

$$\underline{R}_{c_4}^{\mu\eta}(Q_1) = 0.28,\ \underline{R}_{c_4}^{\mu\eta}(Q_2) = 0.28,\ \underline{R}_{c_4}^{\mu\eta}(Q_3) = 0.46,\ \underline{R}_{c_4}^{\mu\eta}(Q_4) = 0.28,$$
$$\underline{R}_{c_4{}'}^{\mu\eta}(Q_1) = 0.19,\ \underline{R}_{c_4{}'}^{\mu\eta}(Q_2) = 0.34,\ \underline{R}_{c_4{}'}^{\mu\eta}(Q_3) = 0.19,\ \underline{R}_{c_4{}'}^{\mu\eta}(Q_4) = 0.35,$$
$$\gamma_{c_4}^{\mu\eta} = 0.32 > \gamma_{c_4{}'}^{\mu\eta} = 0.27.$$

After introducing the tolerance factor, $\gamma_{c_4}^{\mu\eta}$ and $\gamma_{c_4{}'}^{\mu\eta}$ are no longer equal. According to Example 4, we know that $c_4$ and $c_4{}'$ are different, and they have two decision classes whose attribute values are exchanged. In hierarchical classification, these two decision classes are in different positions. According to common sense, in this case, $c_4$ and $c_4{}'$ should have different dependencies. It is evident that introduction of the tolerance index successfully described the influence of connections between decision classes on attribute dependency $\gamma^{\mu\eta}$. It is easy to calculate that when the hierarchical classification has only one layer, the tolerance index is always 1. At this point, IFRS-II degenerates into IFRS-I, which exactly fits the situation of flat classification. When the hierarchical

**Algorithm 2** Attribute reduction based on improved fuzzy rough set.

**Input:** A decision information system $S = (U, AT \cup D)$;
**Output:** Attribute subset $Red$ obtained after attribute reduction;

$Red = AT$;
**for** each $\alpha \subseteq AT$ **do**
  Calculate $\gamma_{\alpha}^{\mu\eta}$;/*Algorithm 1*/
**end for**
Sort the results in ascending order to $conT\_rank$;
**while** $conT\_rank \neq \emptyset$ **do**
  $c = conT\_rank[1]$;
  $conT\_rank = conT\_rank - \{c\}$;
  Calculate $\gamma_{Red}^{\mu\eta}$ and $\gamma_{Red-\{c\}}^{\mu\eta}$;/*Algorithm 1*/
  $\xi_c = \gamma_{Red}^{\mu\eta} - \gamma_{Red-\{c\}}^{\mu\eta}$;
  **if** $\xi_c < \varepsilon$ **then**
    $Red = Red - \{c\}$;
  **end if**
**end while**
**for** i = 1:length($Red$) **do**
  Use $Red'$ to represent columns 1 to $i$ of $Red$;
  Calculate the classification accuracy of $Red'$;
  $maxAccuracy = max(maxAccuracy, accuracy(Red'))$;
**end for**
Find the subset $Red_{sub}$ corresponding to $maxAccuracy$;
$Red = Red_{sub}$;

classification has more than one layer, the tolerance index changes accordingly, affecting the attribute dependency as well due to the hierarchical structure. Therefore, when hierarchical classification exists in decision classes, this paper successfully applies the model to hierarchical classification through the connection of tolerance index.

From the previous analysis, the fuzzy rough set model that considers both stability differences and hierarchical classification is effective to measure the recognition ability of attribute subsets. The larger the difference between decision classes, the better the recognition ability. Furthermore, since our fuzzy rough set model is based on sample set, it exhibits better robustness compared to fuzzy rough set model based on independent sample.

Suppose $\alpha$ is the conditional attribute of the decision information system, $B \subseteq AT$, then the importance of $\alpha$ to the decision attribute can be defined as:

$$\xi_{\alpha} = \gamma_{B\cup\alpha}^{\mu\eta} - \gamma_{B}^{\mu\eta} \tag{13}$$

Given the utilization of hierarchical classification in this paper, the delineation of decision class hierarchical structure is essential. This paper adopts the utilization of sample quantiles as a method for its construction. For a number set $N = \{x_1, x_2, ..., x_s\}$, the sample quantile of order $p$ is the numerical value $q_p$ that satisfies $p = \frac{\left|\{x_i \leq q_p | i=1,2,...,\upsilon\}\right|}{n}$. All labels of the decision attribute are marked as follows:

$$\begin{aligned}
&l_0 = \{q_0 \leq d(x) \leq q_1\}, \quad l_1 = \{q_0 \leq d(x) \leq q_{0.5}\}, \quad l_2 = \{q_{0.5} \leq d(x) \leq q_1\},\\
&l_3 = \{q_0 \leq d(x) \leq q_{0.25}\}, \quad l_4 = \{q_{0.25} \leq d(x) \leq q_{0.5}\}, \quad l_5 = \{q_{0.5} \leq d(x) \leq q_{0.75}\},\\
&l_6 = \{q_{0.75} \leq d(x) \leq q_1\}, \quad l_7 = \{q_0 \leq d(x) \leq q_{0.125}\}, \quad l_8 = \{q_{0.125} \leq d(x) \leq q_{0.25}\},\\
&l_9 = \{q_{0.25} \leq d(x) \leq q_{0.375}\}, \quad l_{10} = \{q_{0.375} \leq d(x) \leq q_{0.5}\}, \quad l_{11} = \{q_{0.5} \leq d(x) \leq q_{0.625}\},\\
&l_{12} = \{q_{0.625} \leq d(x) \leq q_{0.75}\}, \quad l_{13} = \{q_{0.75} \leq d(x) \leq q_{0.875}\}, \quad l_{14} = \{q_{0.875} \leq d(x) \leq q_1\},
\end{aligned}$$

where $d(x)$ represents the decision attribute value of sample $x$. The following experiments all use the seven tree structures $T_i = \{T_1, T_2, T_3, T_4, T_5, T_6, T_7\}$ defined by the label set $L = \{l_1, l_2, ..., l_{14}\}$ above.

$$\begin{aligned}
&L_1 = \{l_0, l_1, l_2\},\ L_2 = L_1 \cup \{l_3, l_4\},\ L_3 = L_2 \cup \{l_5, l_6\},\ L_4 = L_3 \cup \{l_7, l_8\},\\
&L_5 = L_4 \cup \{l_9, l_{10}\},\ L_6 = L_5 \cup \{l_{10}, l_{11}\},\ L_7 = L.
\end{aligned}$$

The attribute reduction based on IFRS-II (ARIFRS-II) proposed in this paper is shown in Algorithm 2. Firstly, our algorithm calculates the dependency based on the sample set, making it less susceptible to the influence of individual erroneous samples. It has the characteristics of high robustness and strong noise resistance. Subsequent Section 5 will also conduct corresponding experiments on noisy data. Then, our algorithm takes into account the differences in stability between decision classes and hierarchical classification, enhancing our utilization of the data and improving classification accuracy. The time complexity of the algorithm is $O(mn^2)$, where $m$ is the number of conditional attributes and $n$ is the number of decision classes. Since $n$ refers to the number of decision classes rather than the number of samples, the overall time complexity is not high, and the algorithm can be extended in the future by changing the calculation formula of attribute importance $\xi_c$. In addition, $\varepsilon$ is a threshold to control the importance of conditional attribute $c$ and the appropriate number of subsets of attributes can be obtained by adjusting the threshold $\varepsilon$.

**Table 3**
Description of data.

| ID | Dataset | Samples | Attributes | Decision Attribute |
|---|---|---|---|---|
| 1 | Travel reviews | 980 | 10 | 3rd |
| 2 | Dog phenotypes and genotypes [40] | 340 | 11 | 2nd |
| 3 | Wine | 178 | 13 | 13rd |
| 4 | Yavno Fox Data [41] | 256 | 13 | 1st |
| 5 | Arnold et al. 2016 functecol dataset [42] | 290 | 16 | 16th |
| 6 | Dataset of individual values [43] | 1350 | 17 | 12nd |
| 7 | FunctEcol2018 Dataset frigatebirds [44] | 103 | 17 | 6th |
| 8 | Parkinsons | 195 | 23 | 2nd |
| 9 | Microsatellite loci scores [45] | 543 | 24 | 24th |
| 10 | Host biomarkers [46] | 189 | 25 | 14th |
| 11 | Ionosphere | 351 | 34 | 15th |
| 12 | Flow Meter D | 180 | 44 | 26th |
| 13 | Spect | 269 | 45 | 43rd |
| 14 | Cribroconcha honggulelenggenesis [47] | 178 | 77 | 1st |
| 15 | Relative warp scores of upper view [48] | 254 | 121 | 4th |
| 16 | LSVT voice rehabilitation | 126 | 310 | 189th |

**Table 4**
Comparison of attribute reduction algorithms.

| Algorithm | Fuzzy rough set | Hierarchical classification | stability factor |
|---|---|---|---|
| AR-IFRS (proposed) | √ | √ | √ |
| KFRS | √ | √ | × |
| SIB | √ | √ | × |
| HCIE | × (Rough set) | √ | × |

## 5. Experiment and analysis

This section verifies the advantages of our proposed attribute reduction through relevant experiments. The experimental environments are Windows 10, Intel Core (TM) I5-10500 CPU (3.10 GHz) and 16 GB RAM, and the experimental platform is MATLAB 2022a. The databases for this paper are from UCI [39] and Dryad. We select the column with the largest variance as the decision attribute, and then normalize the conditional attribute column of the original dataset. In addition, we removed samples with incomplete attribute values from the dataset, and also removed meaningless conditional attributes, such as serial numbers, leaving only numerical attributes. As shown in Table 3, we listed the information of the experimental datsets.

### *5.1. Experiment on attribute reduction*

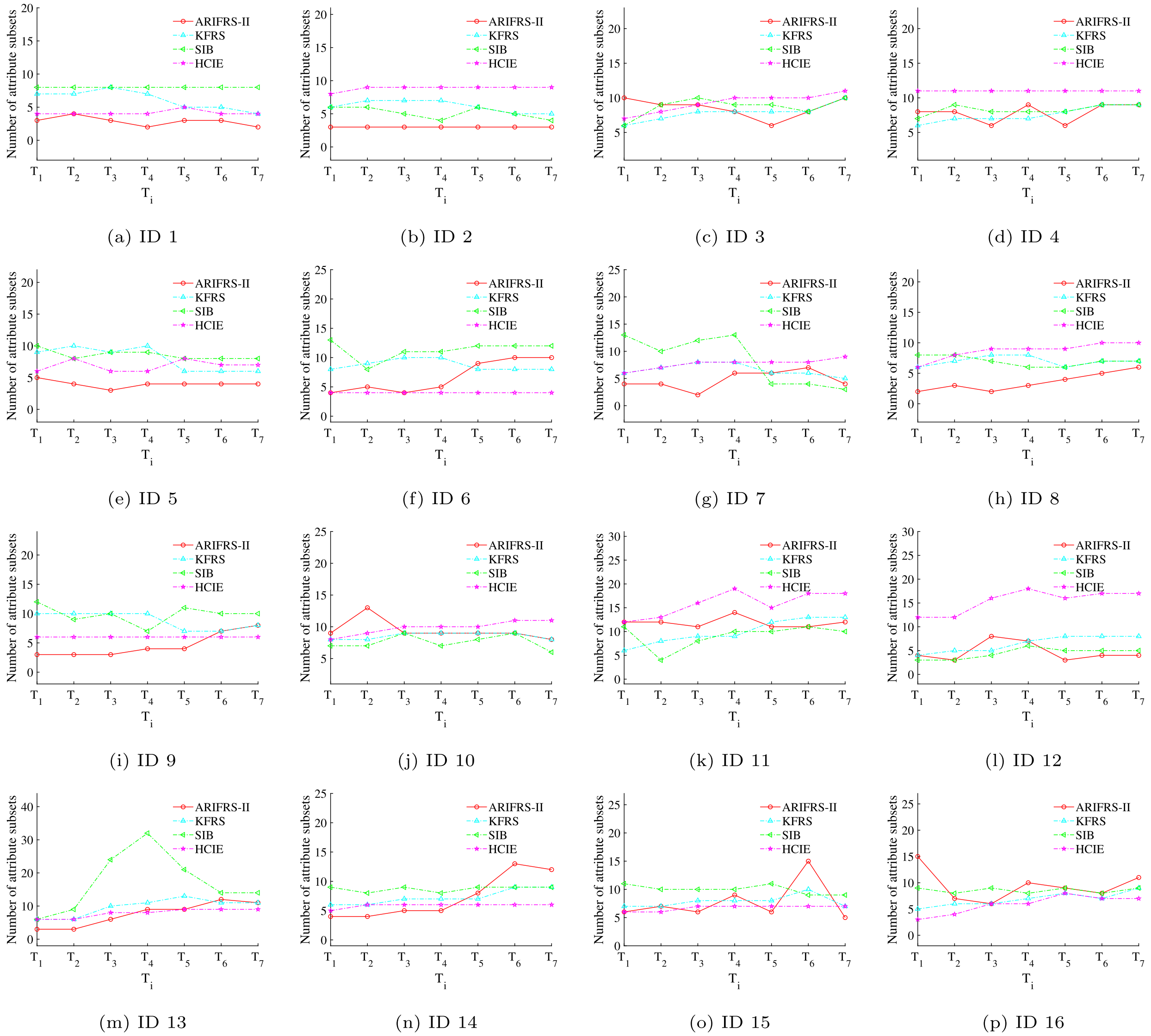
In general, a decision information system may contain redundant attributes, and the goal of attribute reduction is to find the smallest subset of conditional attributes that can preserve the information of the decision information system. This section evaluates the performance of attribute reduction based on improved fuzzy rough set (ARIFRS-II):

(1) In order to better evaluate the performance of ARIFRS-II, we compare 16 datasets with three popular attribute reduction algorithms: hierarchical algorithm based on kernelized fuzzy rough set (KFRS) [49], the sibling search strategy (SIB) [30] and attribute reduction based on the hierarchical conditional information entropy (HCIE) [27]. The comparison of these algorithms is shown in Table 4.

(2) We consider two properties of these algorithms, which are the number of attributes after reduction and the classification accuracy. Firstly, we compare the reduced subsets of our algorithm with the other three algorithms to demonstrate the feasibility of our experiment in reducing the number of subsets. Then two classifiers are used to evaluate the classification accuracy: K-nearest neighbor (KNN, $K = 5$) classifier and Naive Bayes classifier, and each accuracy result is calculated by ten-fold cross-validation. In ARIFRS-II, KNN classifier ($K = 5$) is used to calculate the classification accuracy. $\varepsilon$ is the threshold controlling the importance of attributes in the remaining subsets after reduction. A higher $\varepsilon$ implies higher importance requirements for the remaining subsets, resulting in higher classification accuracy and fewer remaining subsets. Conversely, a smaller $\varepsilon$ leads to lower classification accuracy for the remaining subsets and a larger number of remaining subsets. We conducted multiple experiments by varying the threshold $\varepsilon$ to balance the number of subsets with their importance. Consequently, for the upcoming experiments, we set $\varepsilon$ to 0.01. In future experiments, the value of $\varepsilon$ can be adjusted based on the desired number of subsets and accuracy requirements.

#### *5.1.1. Comparison of the number of attribute reduction subsets*

The following presents a comparison of the subset numbers after attribute reduction using different algorithms across 16 datasets. The serial numbers corresponding to Table 3 replace the dataset names, and it can be seen from Fig. 5 and Table 5:




(a) ID 1 (b) ID 2 (c) ID 3 (d) ID 4

(e) ID 5 (f) ID 6 (g) ID 7 (h) ID 8

(i) ID 9 (j) ID 10 (k) ID 11 (l) ID 12

(m) ID 13 (n) ID 14 (o) ID 15 (p) ID 16

**Fig. 5.** Comparison of the number of subsets after attribute reduction.

(1) For ARIFRS-II, under the seven tree structures of 16 datasets, more than 83.9% of the datasets reduced at least half of the redundant attributes. As the dimension increases, ARIFRS-II can even eliminate more than 95% of redundant attributes, such as dataset ID 16.

(2) In Table 5, the average number of attribute subsets after reduction using seven tree structures is used to reflect the reduction situation of 16 datasets. It is visually observed that the number of attribute subsets after reduction for ARIFRS-II is appropriate and is minimal in the case of seven datasets.

The above analysis shows that ARIFRS-II is reasonable in the number of subsets after attribute reduction.

### 5.1.2. *Comparison of classification accuracy of attribute reduction*

Tables 6-9 and Tables 10-13 illustrate the classification accuracy of 16 datasets using the KNN classifier and Naive Bayes classifier across various attribute reduction algorithms utilizing hierarchical classification. In this paper, $T_1$, $T_2$, $T_3$ and $T_4$ are selected as representatives for analysis, the following observations can be made:

(1) The ARIFRS-II exhibits the significant advantages in both KNN and Naive Bayes classifiers, with the majority of classification accuracy values surpassing those of other algorithms.

(2) The ARIFRS-II algorithm performs well in different tree structures, which indicates the effectiveness of the tolerance index designed in the application of hierarchical classification.

(3) The ARIFRS-II consistently achieves high classification accuracy even with datasets containing a small number of samples. Take ID 7, for instance, which comprises only 103 data samples. Since our algorithm is implemented based on the sample set, it

**Table 5**
The average number of attribute subsets after reduction in seven tree structures.

| Dataset (ID) | ARIFRS-II | KFRS | SIB | HCIE | Raw Data |
|---|---|---|---|---|---|
| 1 | **2.9** | 6.1 | 8.0 | 4.1 | 9.0 |
| 2 | **3.0** | 6.1 | 5.1 | 8.9 | 10.0 |
| 3 | 8.6 | **7.9** | 8.7 | 9.3 | 12.0 |
| 4 | 7.9 | **7.6** | 8.3 | 11.0 | 12.0 |
| 5 | **4.0** | 8.0 | 8.6 | 6.9 | 15.0 |
| 6 | 6.7 | 8.7 | 11.3 | **4.0** | 16.0 |
| 7 | **4.7** | 6.6 | 8.4 | 7.7 | 16.0 |
| 8 | **3.6** | 7.0 | 7.0 | 8.7 | 22.0 |
| 9 | **4.6** | 8.9 | 9.9 | 6.0 | 23.0 |
| 10 | 9.4 | 8.6 | **7.6** | 9.9 | 24.0 |
| 11 | 11.9 | 10.0 | **9.1** | 15.9 | 33.0 |
| 12 | 4.7 | 6.4 | **4.4** | 15.4 | 43.0 |
| 13 | **7.6** | 9.7 | 17.1 | 7.9 | 44.0 |
| 14 | 7.3 | 7.3 | 8.7 | **5.9** | 76.0 |
| 15 | 7.7 | 7.9 | 10.0 | **6.7** | 120.0 |
| 16 | 9.4 | 6.9 | 8.6 | **5.9** | 309.0 |
| Average | **6.5** | 7.7 | 8.8 | 8.4 | 49.0 |

**Table 6**
Classification accuracy (%) under KNN classifier ($K = 5$) ($T_1$).

| Dataset (ID) | ARIFRS-II | KFRS | SIB | HCIE | Raw Data |
|---|---|---|---|---|---|
| 1 | **93.54±0.71** | 92.70±0.81 | 92.00±0.72 | 93.53±0.46 | 92.02±0.81 |
| 2 | **83.94±2.02** | 82.84±3.01 | 82.84±3.01 | 81.78±2.12 | 80.85±2.18 |
| 3 | **88.08±4.24** | 87.89±4.63 | 87.21±3.79 | 86.83±4.14 | 87.86±4.44 |
| 4 | **96.62±0.46** | 96.27±0.88 | 94.64±1.26 | 95.18±1.23 | 95.57±1.03 |
| 5 | **81.03±2.96** | 79.62±2.58 | 78.68±2.14 | 79.29±2.78 | 79.71±2.76 |
| 6 | **97.10±0.54** | 96.62±0.52 | 95.27±0.69 | 95.08±0.47 | 94.92±0.75 |
| 7 | **84.65±5.51** | 82.15±5.15 | 81.66±4.79 | 83.83±5.18 | 84.11±5.30 |
| 8 | **88.91±2.80** | 84.30±2.91 | 85.12±3.24 | 87.17±3.73 | 80.96±3.31 |
| 9 | **80.39±2.65** | 79.49±1.60 | 79.49±1.81 | 80.37±2.46 | 79.49±1.83 |
| 10 | **84.28±2.34** | 80.95±2.32 | 78.69±1.58 | 78.45±2.35 | 81.19±2.50 |
| 11 | **95.46±1.55** | 93.05±1.94 | 94.92±1.03 | 94.34±1.74 | 94.42±1.90 |
| 12 | **94.94±2.28** | 89.75±0.80 | 85.43±2.22 | 87.90±2.81 | 87.65±2.37 |
| 13 | **95.08±2.10** | 94.55±2.07 | 93.72±2.45 | 93.87±2.01 | 91.99±1.82 |
| 14 | **88.71±3.85** | 84.49±2.63 | 84.46±3.02 | 87.86±2.76 | 80.37±1.98 |
| 15 | **83.68±2.14** | 81.58±2.69 | 82.09±1.63 | 82.19±1.79 | 76.32±2.28 |
| 16 | **95.28±2.92** | 92.37±4.62 | 83.05±4.22 | 87.43±5.54 | 88.84±5.70 |
| Average | **89.48±2.44** | 87.42±2.45 | 86.20±2.35 | 87.20±2.60 | 86.02±2.56 |

**Table 7**
Classification accuracy (%) under KNN classifier ($K = 5$) ($T_2$).

| Dataset (ID) | ARIFRS-II | KFRS | SIB | HCIE | Raw Data |
|---|---|---|---|---|---|
| 1 | **84.46±0.83** | 83.25±1.04 | 77.46±1.93 | 84.24±1.26 | 82.72±1.41 |
| 2 | **77.66±1.49** | 75.74±1.78 | 77.00±1.93 | 74.52±2.66 | 74.24±2.71 |
| 3 | **81.70±3.22** | 80.46±2.78 | 80.61±3.37 | 81.26±3.16 | 81.54±3.57 |
| 4 | **94.24±1.28** | 89.20±2.13 | 91.17±1.99 | 92.08±1.73 | 92.55±1.59 |
| 5 | **76.29±4.31** | 75.47±2.05 | 75.35±3.63 | 75.47±4.72 | 75.79±3.35 |
| 6 | **96.35±0.69** | 93.70±0.84 | 91.46±0.63 | 92.77±0.52 | 91.95±0.64 |
| 7 | **75.92±2.25** | 74.20±2.65 | 74.43±4.13 | 75.79±3.74 | 75.54±1.92 |
| 8 | **80.92±4.64** | 79.01±4.29 | 78.99±3.90 | 79.82±3.56 | 76.36±3.42 |
| 9 | **76.19±1.78** | 74.13±1.61 | 73.16±1.97 | 74.42±2.14 | 73.77±1.66 |
| 10 | 76.31±2.02 | **76.39±2.89** | 72.65±4.33 | 74.67±3.21 | 76.10±2.37 |
| 11 | **89.55±2.42** | 88.36±2.65 | 86.32±3.01 | 87.66±2.67 | 87.38±3.90 |
| 12 | **85.61±3.35** | 85.19±7.32 | 80.80±6.70 | 84.66±4.43 | 81.91±7.85 |
| 13 | **92.06±2.11** | 90.94±1.45 | 90.90±1.34 | 91.00±1.55 | 90.47±1.14 |
| 14 | **78.41±5.25** | 77.46±2.75 | 76.81±2.31 | 78.01±4.90 | 74.51±2.80 |
| 15 | **76.82±2.95** | 70.00±2.87 | 70.41±2.76 | 76.56±2.11 | 69.40±6.21 |
| 16 | **86.72±5.96** | 74.82±6.39 | 76.25±4.80 | 85.71±5.89 | 82.39±7.48 |
| Average | **83.08±2.78** | 80.52±2.84 | 79.61±3.05 | 81.79±3.01 | 80.41±3.25 |

**Table 8**
Classification accuracy (%) under KNN classifier ($K = 5$) ($T_3$).

| Dataset (ID) | ARIFRS-II | KFRS | SIB | HCIE | Raw Data |
|---|---|---|---|---|---|
| 1 | **80.91±1.34** | 79.51±1.28 | 79.51±1.28 | 80.36±1.52 | 79.65±1.30 |
| 2 | **72.85±3.39** | 70.70±2.84 | 71.61±2.75 | 68.83±2.44 | 68.71±2.82 |
| 3 | **73.36±6.47** | 72.49±5.86 | 73.15±6.77 | 72.20±6.99 | 72.92±6.62 |
| 4 | **90.81±2.55** | 86.92±4.43 | 87.45±5.05 | 88.18±5.15 | 88.58±5.16 |
| 5 | **67.73±2.03** | 64.32±2.95 | 64.50±2.32 | 65.08±2.46 | 65.52±2.89 |
| 6 | **93.76±0.51** | 91.22±1.05 | 85.14±0.90 | 89.27±0.92 | 88.61±1.33 |
| 7 | **64.99±6.00** | 63.58±5.40 | 62.13±4.84 | 62.56±4.60 | 63.98±6.17 |
| 8 | **75.48±3.80** | 72.60±4.36 | 72.26±4.63 | 73.77±3.87 | 66.71±3.70 |
| 9 | **73.42±2.53** | 70.99±3.29 | 70.18±3.73 | 71.98±2.91 | 70.22±3.50 |
| 10 | **66.39±2.48** | 62.39±3.86 | 63.16±4.27 | 63.24±3.85 | 63.53±4.19 |
| 11 | **84.47±2.69** | 82.04±2.73 | 82.19±3.11 | 82.82±3.11 | 81.24±4.39 |
| 12 | **75.12±3.38** | 74.47±3.47 | 70.33±4.09 | 71.91±6.24 | 71.44±4.97 |
| 13 | **81.25±2.16** | 78.94±2.25 | 76.81±1.15 | 78.95±1.67 | 78.07±1.81 |
| 14 | **69.76±3.15** | 68.81±3.21 | 67.83±2.43 | 69.59±4.21 | 63.48±2.33 |
| 15 | **67.17±2.43** | 64.85±2.63 | 57.19±1.53 | 64.93±3.09 | 60.76±2.09 |
| 16 | **84.90±9.20** | 69.28±5.04 | 63.88±4.28 | 72.13±6.38 | 71.52±5.06 |
| Average | **76.40±3.38** | 73.32±3.41 | 71.71±3.32 | 73.49±3.71 | 72.18±3.65 |

**Table 9**
Classification accuracy (%) under KNN classifier ($K = 5$) ($T_4$).

| Dataset (ID) | ARIFRS-II | KFRS | SIB | HCIE | Raw Data |
|---|---|---|---|---|---|
| 1 | 76.07±1.15 | 74.49±1.19 | 74.35±1.24 | **76.56±1.63** | 74.58±1.07 |
| 2 | **70.47±2.46** | 69.29±1.50 | 70.32±2.37 | 66.66±1.59 | 66.30±1.50 |
| 3 | **73.62±5.15** | 71.91±4.20 | 72.82±5.39 | 71.13±4.78 | 72.52±5.18 |
| 4 | **87.84±4.46** | 85.64±4.20 | 85.44±4.18 | 86.75±4.40 | 87.09±4.56 |
| 5 | **65.12±2.77** | 63.54±1.74 | 63.34±1.55 | 62.76±2.07 | 63.58±2.44 |
| 6 | **92.37±0.65** | 90.60±0.71 | 87.85±0.83 | 91.08±0.59 | 88.51±1.03 |
| 7 | **62.23±5.31** | 59.07±4.75 | 60.05±4.80 | 59.10±4.23 | 61.55±5.13 |
| 8 | 68.76±4.16 | 68.19±4.12 | 68.32±3.63 | **69.33±2.06** | 64.40±2.96 |
| 9 | **73.22±2.82** | 71.03±1.56 | 69.04±1.93 | 71.61±2.53 | 70.28±1.64 |
| 10 | **63.06±4.51** | 61.63±2.51 | 60.43±3.81 | 63.01±2.90 | 62.68±2.74 |
| 11 | **84.07±2.61** | 77.16±3.12 | 79.29±3.30 | 82.48±3.58 | 81.91±4.30 |
| 12 | **68.33±4.18** | 67.40±5.78 | 65.15±3.77 | 66.83±5.04 | 67.93±4.90 |
| 13 | **80.78±2.50** | 79.04±1.90 | 76.81±2.29 | 79.05±2.52 | 77.69±2.47 |
| 14 | **67.24±5.28** | 65.33±2.21 | 65.43±4.15 | 64.55±3.38 | 60.69±2.29 |
| 15 | **64.02±2.37** | 56.80±1.74 | 56.63±2.41 | 63.10±2.22 | 57.89±1.55 |
| 16 | **81.15±5.67** | 67.97±4.78 | 65.76±4.15 | 72.35±4.64 | 72.17±6.98 |
| Average | **73.65±3.50** | 70.57±2.88 | 70.06±3.11 | 71.65±3.01 | 70.61±3.17 |

**Table 10**
Classification accuracy (%) under Naive Bayes classifier ($T_1$).

| Dataset (ID) | ARIFRS-II | KFRS | SIB | HCIE | Raw Data |
|---|---|---|---|---|---|
| 1 | **91.59±1.43** | 86.56±1.20 | 86.81±1.23 | 89.44±1.11 | 86.29±1.33 |
| 2 | **91.14±1.07** | 89.79±1.73 | 89.79±1.73 | 88.79±2.28 | 87.06±2.56 |
| 3 | **89.39±1.62** | 85.93±2.66 | 84.96±2.76 | 86.86±2.33 | 88.76±2.15 |
| 4 | 95.17±1.30 | 94.58±1.34 | 95.06±1.15 | **95.69±1.39** | 95.54±1.53 |
| 5 | **80.46±2.08** | 79.58±1.19 | 78.43±1.41 | 79.39±1.62 | 79.77±1.92 |
| 6 | **96.22±0.43** | 95.94±0.54 | 94.31±0.38 | 93.23±0.43 | 93.16±0.39 |
| 7 | **87.17±3.53** | 86.12±2.72 | 84.54±2.63 | 86.42±2.49 | 85.62±2.41 |
| 8 | **90.85±1.07** | 90.38±1.51 | 89.01±1.88 | 88.99±1.82 | 84.87±4.02 |
| 9 | **80.18±2.30** | 78.02±1.65 | 79.15±1.78 | 79.51±1.51 | 78.88±2.56 |
| 10 | 83.25±1.81 | 83.33±1.98 | 83.04±2.27 | 84.42±1.95 | **84.51±1.28** |
| 11 | **93.76±1.35** | 91.67±1.64 | 93.62±1.45 | 93.70±0.93 | 93.30±1.14 |
| 12 | **89.54±0.35** | 88.52±1.45 | 87.01±1.67 | 86.94±1.72 | 89.14±1.57 |
| 13 | 94.30±1.74 | **95.27±0.83** | 92.94±0.98 | 93.97±1.00 | 93.62±0.66 |
| 14 | 90.51±1.97 | 88.86±1.16 | 87.30±1.90 | **90.76±2.15** | 80.03±1.08 |
| 15 | 83.14±1.65 | 79.46±2.46 | 78.44±1.69 | **83.68±1.19** | 77.40±1.46 |
| 16 | **97.40±1.93** | 95.69±1.96 | 84.53±2.72 | 94.80±2.71 | 95.68±2.24 |
| Average | **89.63±1.60** | 88.11±1.63 | 86.81±1.73 | 88.54±1.66 | 87.10±1.77 |

**Table 11**
Classification accuracy (%) under Naive Bayes classifier ($T_2$).

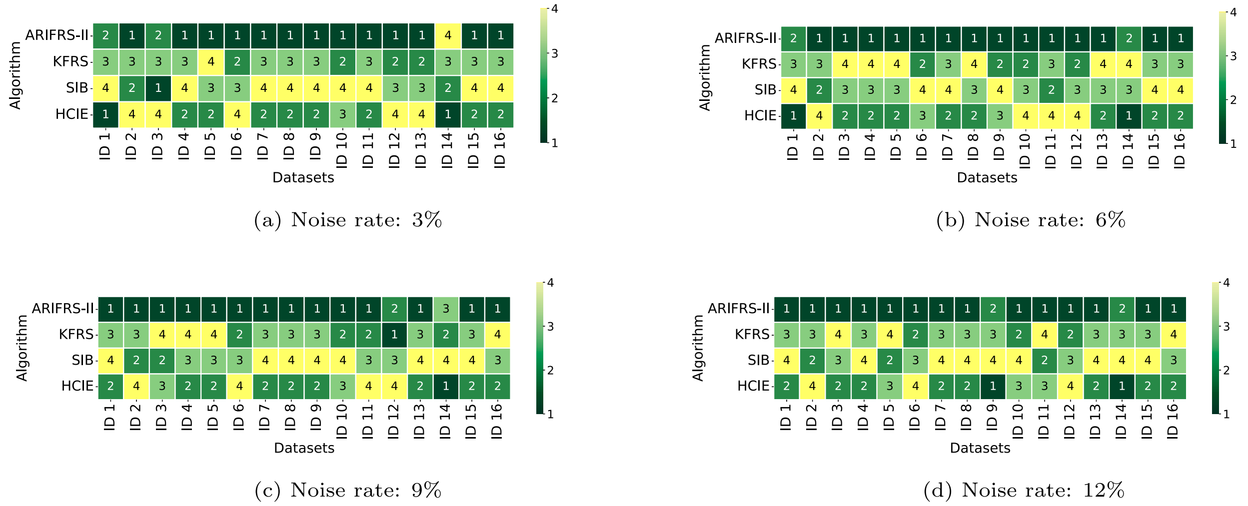
| Dataset (ID) | ARIFRS-II | KFRS | SIB | HCIE | Raw Data |
|---|---|---|---|---|---|
| 1 | **82.85±0.84** | 81.10±1.11 | 72.46±1.40 | 81.17±0.93 | 79.91±1.12 |
| 2 | **84.66±2.25** | 81.79±1.98 | 82.87±2.75 | 79.83±2.87 | 78.59±2.79 |
| 3 | **81.33±2.43** | 78.94±3.64 | 79.45±1.52 | 80.74±2.78 | 80.83±2.36 |
| 4 | **90.34±2.08** | 88.28±1.54 | 89.37±1.77 | 89.93±1.96 | 89.97±2.01 |
| 5 | 68.03±2.99 | 67.13±2.07 | **68.56±2.47** | 67.79±2.21 | 67.29±2.53 |
| 6 | **92.71±0.25** | 91.49±0.70 | 88.95±0.93 | 87.64±0.91 | 88.89±0.71 |
| 7 | **79.76±5.15** | 77.39±3.12 | 75.63±4.86 | 75.91±4.61 | 73.95±3.56 |
| 8 | **81.57±2.81** | 77.11±3.27 | 76.62±3.54 | 75.75±2.88 | 73.11±4.28 |
| 9 | **75.21±1.64** | 73.16±1.24 | 74.37±1.10 | 74.26±1.46 | 73.06±1.43 |
| 10 | 76.42±3.87 | 76.18±4.08 | 75.39±3.62 | 76.24±2.63 | **76.71±3.16** |
| 11 | **89.71±1.39** | 89.60±0.59 | 81.86±1.76 | 88.54±1.51 | 89.24±1.38 |
| 12 | 84.22±5.47 | 86.85±1.75 | **88.19±0.88** | 83.27±2.61 | 86.39±2.51 |
| 13 | 89.30±2.25 | 88.78±1.74 | 89.26±1.51 | 88.31±1.95 | **90.07±1.65** |
| 14 | **79.91±4.11** | 75.82±1.87 | 75.81±2.11 | 76.19±2.44 | 70.32±4.36 |
| 15 | 73.30±2.42 | 64.54±1.95 | 65.53±1.98 | **74.92±1.71** | 69.29±2.83 |
| 16 | **95.15±3.67** | 79.81±4.47 | 75.68±2.75 | 90.55±3.56 | 89.13±4.59 |
| Average | **82.78±2.73** | 79.87±2.19 | 78.75±2.18 | 80.69±2.31 | 79.80±2.58 |

**Table 12**
Classification accuracy (%) under Naive Bayes classifier ($T_3$).

| Dataset (ID) | ARIFRS-II | KFRS | SIB | HCIE | Raw Data |
|---|---|---|---|---|---|
| 1 | **78.54±1.89** | 73.91±0.99 | 73.91±0.99 | 75.46±1.50 | 72.74±1.17 |
| 2 | **82.57±2.11** | 76.95±2.01 | 78.53±2.67 | 74.81±2.42 | 73.20±2.61 |
| 3 | 73.89±3.53 | 72.79±3.59 | 73.14±2.78 | 73.23±3.12 | **73.95±3.22** |
| 4 | 90.36±1.64 | 88.57±1.09 | 90.06±1.58 | 91.62±1.58 | **91.68±1.37** |
| 5 | **64.00±1.98** | 61.47±1.92 | 62.59±1.45 | 62.86±1.80 | 63.72±1.89 |
| 6 | **91.76±0.50** | 89.69±0.51 | 82.40±0.72 | 86.10±0.73 | 86.15±0.38 |
| 7 | 67.11±3.12 | **68.24±3.14** | 66.10±2.47 | 68.10±2.10 | 67.45±2.14 |
| 8 | **80.22±1.89** | 76.99±2.04 | 76.94±2.40 | 76.12±2.55 | 70.16±2.66 |
| 9 | **72.11±1.80** | 68.77±2.29 | 69.81±2.08 | 70.48±2.28 | 69.24±1.85 |
| 10 | 66.94±2.24 | 66.00±2.88 | 65.88±1.92 | **69.70±2.54** | 67.84±1.75 |
| 11 | **85.35±0.93** | 83.58±1.28 | 83.43±2.24 | 84.79±1.85 | 84.33±1.99 |
| 12 | **78.00±2.43** | 75.31±3.03 | 74.36±2.68 | 77.53±3.16 | 76.60±3.53 |
| 13 | **84.39±1.91** | 79.98±2.15 | 77.52±1.55 | 79.65±2.14 | 81.29±1.59 |
| 14 | **76.66±2.53** | 75.01±2.26 | 69.87±2.27 | 74.59±2.56 | 64.94±1.70 |
| 15 | **67.74±1.93** | 62.84±1.53 | 58.64±1.35 | 66.84±2.40 | 59.84±1.15 |
| 16 | **90.56±1.75** | 80.39±3.10 | 68.22±3.69 | 83.45±3.20 | 85.15±5.39 |
| Average | **78.14±2.01** | 75.03±2.11 | 73.21±2.05 | 75.96±2.25 | 74.27±2.15 |

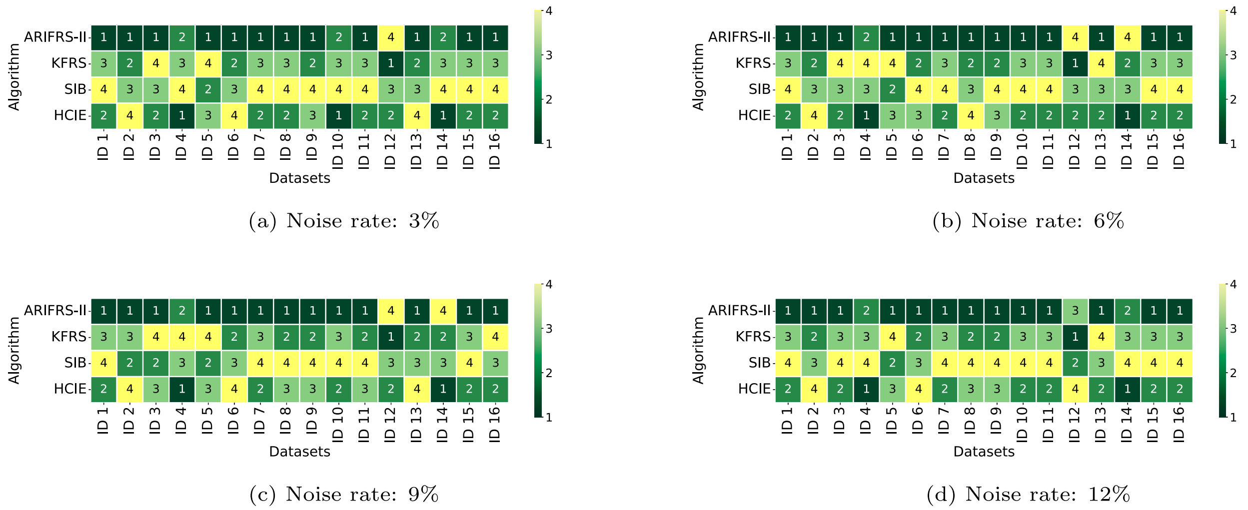
**Table 13**
Classification accuracy (%) under Naive Bayes classifier ($T_4$).

| Dataset (ID) | ARIFRS-II | KFRS | SIB | HCIE | Raw Data |
|---|---|---|---|---|---|
| 1 | **73.19±1.59** | 71.91±1.41 | 71.26±1.39 | 71.84±1.46 | 71.00±1.11 |
| 2 | **76.24±2.39** | 72.60±2.76 | 75.27±2.31 | 69.46±3.30 | 68.00±2.81 |
| 3 | **70.93±2.86** | 68.43±2.28 | 68.51±3.47 | 69.16±3.14 | 69.84±2.90 |
| 4 | **88.90±1.83** | 87.68±1.61 | 87.00±2.17 | 88.22±2.06 | 88.89±2.10 |
| 5 | 58.24±4.25 | 56.88±1.99 | 57.22±2.07 | **58.90±2.63** | 58.29±2.76 |
| 6 | **88.59±0.96** | 87.04±0.64 | 83.67±0.86 | 86.14±0.60 | 84.07±0.92 |
| 7 | **68.01±3.70** | 63.74±4.17 | 62.18±3.39 | 63.82±3.56 | 62.43±2.93 |
| 8 | **70.72±3.22** | 68.17±2.92 | 67.98±3.15 | 67.20±3.93 | 63.12±3.77 |
| 9 | 68.46±2.71 | **68.86±1.88** | 62.36±1.80 | 67.78±2.44 | 68.49±2.37 |
| 10 | **64.93±1.68** | 61.43±4.36 | 61.43±2.05 | 64.87±3.66 | 64.37±2.51 |
| 11 | **83.12±1.15** | 73.88±2.72 | 77.22±2.60 | 82.27±1.76 | 82.12±1.42 |
| 12 | 72.87±3.02 | 73.58±2.82 | 73.89±2.58 | **76.87±3.64** | 73.33±3.76 |
| 13 | **82.70±1.36** | 79.20±2.48 | 77.04±1.58 | 78.21±2.83 | 79.27±1.56 |
| 14 | **74.35±3.45** | 70.51±2.82 | 66.52±3.01 | 72.33±2.55 | 61.32±1.44 |
| 15 | **62.98±3.30** | 56.69±2.07 | 55.67±1.72 | 62.27±2.65 | 58.22±1.13 |
| 16 | **88.40±2.92** | 76.82±2.51 | 66.44±2.23 | 81.79±1.55 | 85.52±3.65 |
| Average | **74.54±2.53** | 71.09±2.47 | 69.60±2.27 | 72.57±2.61 | 71.14±2.32 |




(a) Noise rate: 3%
(b) Noise rate: 6%
(c) Noise rate: 9%
(d) Noise rate: 12%

**Fig. 6.** Comparison on ranking of classification accuracy (KNN classifier ($K = 5$)) under each noise rate.




(a) Noise rate: 3%
(b) Noise rate: 6%
(c) Noise rate: 9%
(d) Noise rate: 12%

**Fig. 7.** Comparison on ranking of classification accuracy (Naive Bayes classifier) under each noise rate.

can increase the robustness against noise data by processing a sample set as an information granule from the perspective of granular computing. Therefore, the presence of a few erroneous samples does not significantly disrupt the outcome. For this reason, we set up a noise experiment and construct a dataset with 3%, 6%, 9%, and 12% noise. Specifically, after dividing the decision classes using sample quantiles, randomly select 3%, 6%, 9%, and 12% samples and change their decision class labels. Finally, the average classification accuracy (KNN classifier and Naive Bayes classifier) of noisy datasets on seven decision trees is compared under four attribute reduction algorithms. As shown in Fig. 6 and Fig. 7, the numbers 1, 2, 3, and 4 on the figure represent the ranking of classification accuracy. For example, on the ID 2 with a noise ratio of 6%, ARIFRS-II has the highest classification accuracy, marked as 1. It can be seen that in most cases, the classification accuracy of ARIFRS-II is the highest. Furthermore, the range of conditional attributes used in our experiments spans from ten to over one hundred, all exhibiting high classification accuracy post-reduction. This highlights the suitability of ARIFRS-II for datasets of diverse sizes.

Table 14 and Table 15 present the increase in classification accuracy before and after ARIFRS-II, which are calculated using the following formula:

$$Incr = \frac{CA(ARIFRS - II) - CA(Raw)}{CA(Raw)} \times 100 \tag{14}$$

where $CA(ARIFRS - II)$ and $CA(Raw)$ represent the classification accuracy of the remaining attribute after ARIFRS-II reduction and the classification accuracy of the raw attribute, respectively.

**Table 14**
Increment (%) of classification accuracy before and after ARIFRS-II (KNN classifier ($K = 5$)).

| Dataset (ID) | $T_1$ | $T_2$ | $T_3$ | $T_4$ |
|---|---|---|---|---|
| 1 | +1.64 | +2.10 | +1.58 | +2.00 |
| 2 | +3.82 | +4.61 | +6.02 | +6.28 |
| 3 | +0.25 | +0.19 | +0.60 | +1.51 |
| 4 | +1.09 | +1.83 | +2.52 | +0.87 |
| 5 | +1.66 | +0.67 | +3.37 | +2.43 |
| 6 | +2.30 | +4.78 | +5.81 | +4.36 |
| 7 | +0.65 | +0.50 | +1.59 | +1.09 |
| 8 | +9.82 | +5.97 | +13.16 | +6.77 |
| 9 | +1.13 | +3.28 | +4.55 | +4.19 |
| 10 | +3.80 | +0.27 | +4.50 | +0.62 |
| 11 | +1.11 | +2.49 | +3.98 | +2.64 |
| 12 | +8.31 | +4.51 | +5.16 | +0.59 |
| 13 | +3.37 | +1.76 | +4.08 | +3.97 |
| 14 | +10.38 | +5.24 | +9.89 | +10.79 |
| 15 | +9.64 | +10.70 | +10.55 | +10.60 |
| 16 | +7.25 | +5.25 | +18.71 | +12.44 |

**Table 15**
Increment (%) of classification accuracy before and after ARIFRS-II (Naive Bayes classifier).

| Dataset (ID) | $T_1$ | $T_2$ | $T_3$ | $T_4$ |
|---|---|---|---|---|
| 1 | +6.14 | +3.68 | +7.98 | +3.08 |
| 2 | +4.69 | +7.73 | +12.80 | +12.12 |
| 3 | +0.71 | +0.62 | -0.08 | +1.56 |
| 4 | -0.39 | +0.42 | -1.44 | +0.01 |
| 5 | +0.86 | +1.11 | +0.44 | -0.08 |
| 6 | +3.28 | +4.29 | +6.51 | +5.38 |
| 7 | +1.80 | +7.86 | -0.51 | +8.93 |
| 8 | +7.05 | +11.57 | +14.35 | +12.03 |
| 9 | +1.65 | +2.94 | +4.13 | -0.05 |
| 10 | -1.50 | -0.37 | -1.33 | +0.87 |
| 11 | +0.49 | +0.52 | +1.21 | +1.22 |
| 12 | +0.45 | -2.51 | +1.83 | -0.64 |
| 13 | +0.73 | -0.85 | +3.81 | +4.33 |
| 14 | +13.11 | +13.63 | +18.04 | +21.25 |
| 15 | +7.41 | +5.78 | +13.20 | +8.17 |
| 16 | +1.79 | +6.76 | +6.36 | +3.37 |

**Table 16**
Friedman test results of four attribute reduction methods under KNN ($K = 5$) and Naive Bayes classifiers.

| Granularity Layer | Classifier | Mean Ranking | | | | $p_{\chi^2}$ | $p_F$ |
|---|---|---|---|---|---|---|---|
| | | ARIFRS-II | KFRS | SIB | HCIE | | |
| $T_1$ | KNN ($K = 5$) | 1.00 | 2.69 | 3.44 | 2.88 | 5.98 E-07 | 1.21 E-10 |
| | Naive Bayes | 1.38 | 2.69 | 3.50 | 2.44 | 6.14 E-05 | 3.47 E-06 |
| $T_2$ | KNN ($K = 5$) | 1.06 | 2.94 | 3.56 | 2.44 | 4.01 E-07 | 3.79 E-11 |
| | Naive Bayes | 1.25 | 3.00 | 2.94 | 2.81 | 1.56 E-04 | 1.70 E-05 |
| $T_3$ | KNN ($K = 5$) | 1.00 | 3.06 | 3.50 | 2.44 | 1.73 E-07 | 2.71 E-12 |
| | Naive Bayes | 1.25 | 2.88 | 3.56 | 2.31 | 4.57 E-06 | 1.98 E-08 |
| $T_4$ | KNN ($K = 5$) | 1.13 | 3.06 | 3.31 | 2.50 | 4.57 E-06 | 1.98 E-08 |
| | Naive Bayes | 1.31 | 2.88 | 3.44 | 2.38 | 3.21 E-05 | 1.07 E-06 |

The results indicate an enhancement in classification accuracy of all datasets under the KNN classifier. Moreover, under Naive Bayes classifiers, 81.3% of datasets exhibited improved classification accuracy. Therefore, ARIFRS-II effectively eliminates redundant data from the raw data.

In order to make the above experiments more convincing, as shown in Table 16, this paper calculates the Friedman test results of four attribute reduction algorithms on 16 datasets, where $p_{\chi^2}$ and $p_F$ represent $p$-value statistics. It can be seen from the table that the $p$-value$\in [10^{-4}, 10^{-12}]$, that is, there are significant differences between these four algorithms in these datasets. Fig. 8 shows the results of Bonferroni-Dunn test, and the colored lines in the figure indicate that there is no significant difference. ARIFRS-II algorithm has obvious differences and better performance compared with KFRS, SIB, and HCIE. Table 17 lists the $p$-values in the

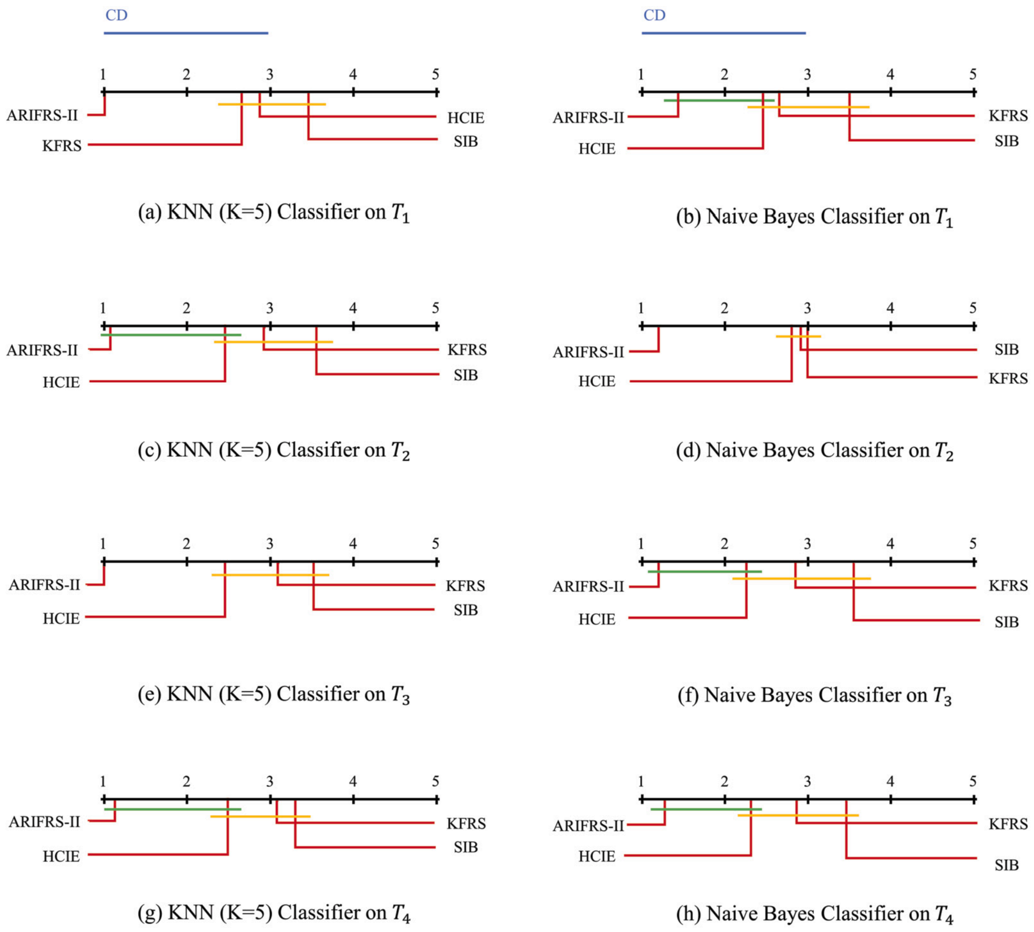



**Fig. 8.** Visualization of post hoc tests for the data from Table 16.

**Table 17**
Wilcoxon signed rank test results between ARIFRS-II and other attribute reduction algorithms under KNN ($K = 5$) and Naive Bayes classifiers.

| Granularity Layer | Classifier | $p_u$ | | |
|---|---|---|---|---|
| | | KFRS | SIB | HCIE |
| $T_1$ | KNN ($K = 5$) | 4.38 E-04 | 4.38 E-04 | 4.38 E-04 |
| | Naive Bayes | 1.61 E-03 | 4.38 E-04 | 1.31 E-02 |
| $T_2$ | KNN ($K = 5$) | 5.31 E-04 | 4.38 E-04 | 4.38 E-04 |
| | Naive Bayes | 3.20 E-03 | 3.20 E-03 | 2.28 E-03 |
| $T_3$ | KNN ($K = 5$) | 4.38 E-04 | 4.38 E-04 | 4.38 E-04 |
| | Naive Bayes | 7.76 E-04 | 4.38 E-04 | 1.74 E-02 |
| $T_4$ | KNN ($K = 5$) | 4.38 E-04 | 4.38 E-04 | 1.12 E-03 |
| | Naive Bayes | 7.76 E-04 | 6.43 E-04 | 5.23 E-03 |

Wilcoxon signed rank test between ARIFRS-II and the other three algorithms. From the smaller value range of $p$-value$\in [10^{-2}, 10^{-4}]$, which indicates that the median accuracies of ARIFRS-II are obviously better than that of HCIE.

## 5.2. Parameter analysis

The following experiments are conducted based on formula (11) for calculating the tolerance index $\eta^H$ in hierarchical structure. Let $\eta = \{0.1, 0.2, 0.3, 0.4, 0.5\}$, which denotes the approximation degree between decision classes, and let $Dis^\mu = \{0.1, 0.2, 0.3, 0.4, 0.5\}$, which represents the distance considering the stability factor between the decision classes corresponding to $\eta$. We can observe the impact of $\eta$ and $Dis^\mu$ on $\eta^H$ from Fig. 9.

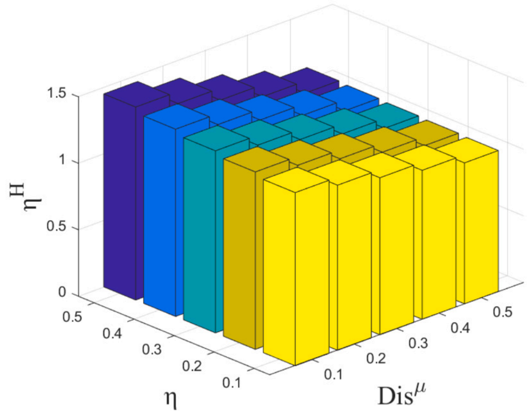



**Fig. 9.** The influence trend of tolerance indexes in hierarchical classification.

**Table 18**
Comparison of classification accuracy (%) (KNN classifier ($K = 5$)) in four cases.

| $T_i$ | AR-None | ARIFRS-I | ARIFRS'-I | ARIFRS-II |
|---|---|---|---|---|
| $T_1$ | 89.32 | **89.48** | 89.32 | **89.48** |
| $T_2$ | 82.70 | 82.84 | 82.92 | **83.08** |
| $T_3$ | 74.74 | 75.71 | 76.30 | **76.40** |
| $T_4$ | 71.19 | 72.27 | 73.59 | **73.65** |

**Table 19**
Comparison of classification accuracy (%) (Naive Bayes classifier) in four cases.

| $T_i$ | AR-None | ARIFRS-I | ARIFRS'-I | ARIFRS-II |
|---|---|---|---|---|
| $T_1$ | **89.63** | 89.47 | 89.47 | **89.63** |
| $T_2$ | 82.03 | 82.75 | 82.20 | **82.78** |
| $T_3$ | 77.83 | 78.03 | 76.53 | **78.14** |
| $T_4$ | 72.34 | 73.91 | 70.44 | **74.54** |

(1) When $\eta$ is constant, $\eta^H$ decreases with the increase of $Dis^\mu$. This demonstrates that when the approximation degree between decision classes is constant, the increase of $Dis^\mu$ between decision classes reduces the need to consider the hierarchical classification, that is, the $\eta^H$ decreases, and vice versa.

(2) When $Dis^\mu$ is constant, $\eta^H$ increases with the increase of $\eta$. This demonstrates that when $Dis^\mu$ between decision classes is constant, the increase of the approximation degree $\eta$ among decision classes makes the tolerance distance among decision classes larger, that is, the increase of $\eta^H$, and vice versa.

In addition, the practicability of the improved fuzzy rough model proposed in this paper is verified by comparing the changes in classification accuracy before and after the introduction of stability factor and hierarchical classification, respectively. In order to make the conclusion more convincing, 16 datasets in Table 3 are used for comparative analysis, and the average value of classification accuracy of 16 datasets is taken as the comparison result. We carry out four attribute reduction experiments respectively and calculate the classification accuracy, denoted by the symbols AR-None, ARIFRS-I, ARIFRS'-I and ARIFRS-II, respectively. We choose the case of $T_1$ to $T_4$. AR-None denotes the attribute reduction without considering stability differences and hierarchical classification, ARIFRS-I denotes the attribute reduction only considers stability differences, and ARIFRS'-I denotes the attribute reduction that only considers hierarchical classification.

From Table 18 and Table 19, we can observe that the deeper the hierarchical structure, the more significant the role of ARIFRS-II, which also indicates the advantage of ARIFRS-II in dealing with hierarchical classification. Moreover, compared to before introducing stability factor and hierarchical classification, ARIFRS-II has significantly improved classification accuracy under both KNN and Naive Bayes classifiers except for the scenario involving tree structure $T_1$. With only two decision classes in tree structure $T_1$, the impact of the stability factor and hierarchical classification is small, leading to very similar classification accuracy across the four cases.

## 6. Conclusion

Attribute reduction, as a crucial step in data preprocessing, plays a significant role in enhancing efficiency and interpretability of classification models. This paper introduced two types of fuzzy rough set (IFRS-I and IFRS-II) based on sample set (divided by decision classes), which enhanced the robustness against noise data. Subsequently, a stable and effective attribute reduction based on IFRS-

II, named ARIFRS-II, was constructed for hierarchical classification. Our attribute reduction algorithm not only considered stability information other than numerical values but also could be applied to hierarchical classification. Moreover, due to its sample set based characteristics, its robustness was also enhanced. Experimental outcomes indicated that compared to existing popular algorithms, ARIFRS-II outperformed the current popular algorithms in most cases in terms of classification accuracy and robustness. Furthermore, we conducted experiments comparing our algorithm with and without stability information and hierarchical classification. The experimental results confirmed the superiority of ARIFRS-II.

From the perspective of multi-granularity, the input of current attribute reduction methods is mainly based on the finest-grained point, which affects their efficiency to some extent. Although the Pawlak rough set are represented by equivalence classes and have a good knowledge representation capability, it failed to automatically process continuous data; the fuzzy rough set and neighborhood rough set are able to process continuous data, but their information granules comprise of sample points rather than equivalence classes, and thus lose the ability to represent knowledge. Therefore, our model is exploratory and still has certain limitations, especially the inputs of our model are individual fuzzy information granules, which reduces the decision efficiency and generality.

As the importance of uncertainty and noise-related issues grows, the future direction of attribute reduction may focus more on the directions as follows:

(1) By introducing granular ball computing to hierarchical classification based on fuzzy rough set, more robust and efficient attribute reduction can be constructed. For example, when faced with a large volume of data samples, utilizing granular ball sampling and then incorporating it into hierarchical classification within the framework of fuzzy rough set is an effective approach.

(2) To obtain the objective and effective reduction thresholds, it is beneficial to introduce uncertainty into three-way attribute reduction with fuzzy rough set.

(3) An optimal decision space selection mechanism will be designed for hierarchical classification by considering the user requirements for decision precision and computational expense, respectively.

These works will be an enrichment to fuzzy rough set from a new perspective and contribute to the progress of attribute reduction.

## CRediT authorship contribution statement

**Jie Yang:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization. **Xiaodan Qin:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Guoyin Wang:** Writing – review & editing, Supervision. **Qinghua Zhang:** Writing – review & editing, Supervision. **Shuai Li:** Writing – review & editing, Supervision. **Di Wu:** Writing – review & editing, Validation, Supervision, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgements

This work was supported by the National Science Foundation of China (Grant number 62066049, 62176070, 62366033), the Guizhou Provincial Department of Education Colleges and Universities Science and Technology Innovation Team (QJJ [2023] 084), Excellent Young Scientific and Technological Talents Foundation of Guizhou Province (QKH-platform talent (2021) 5627), Science and Technology Top Talent Project of Guizhou Education Department (QJJ2022 [088]), Jiangxi Provincial Natural Science Foundation (No. 20232BAB202049), Science and Technology Project of Zunyi (ZSKRPT[2023] 3).

## References

[1] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982) 341–356, https://doi.org/10.1007/BF01001956.

[2] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (2–3) (1990) 191–209, https://doi.org/10.1080/03081079008935107.

[3] K. Zhang, J. Zhan, W.-Z. Wu, Novel fuzzy rough set models and corresponding applications to multi-criteria decision-making, Fuzzy Sets Syst. 383 (2020) 92–126, https://doi.org/10.1016/j.fss.2019.06.019.

[4] J. Ye, J. Zhan, W. Ding, H. Fujita, A novel fuzzy rough set model with fuzzy neighborhood operators, Inf. Sci. 544 (2021) 266–297, https://doi.org/10.1016/j.ins.2020.07.030.

[5] J. Zhan, X. Zhang, Y. Yao, Covering based multigranulation fuzzy rough sets and corresponding applications, Artif. Intell. Rev. 53 (2020) 1093–1126, https://doi.org/10.1007/s10462-019-09690-y.

[6] S.A. Shehab, A. Darwish, A.E. Hassanien, S. R. G. in Egypt, Water quality classification model with small features and class imbalance based on fuzzy rough sets, in: Environment, Development and Sustainability, 2023, pp. 1–19.

[7] F. Liu, A.A. Sekh, C. Quek, G.S. Ng, D.K. Prasad, Rs-herr: a rough set-based Hebbian rule reduction neuro-fuzzy system, Neural Comput. Appl. 33 (2021) 1123–1137, https://doi.org/10.1007/s00521-020-04997-2.

[8] B. Sun, W. Ma, X. Chen, Variable precision multigranulation rough fuzzy set approach to multiple attribute group decision-making based on $\lambda$-similarity relation, Comput. Ind. Eng. 127 (2019) 326–343, https://doi.org/10.1016/j.cie.2018.10.009.

[9] B. Sun, C. Qi, W. Ma, T. Wang, L. Zhang, C. Jiang, Variable precision diversified attribute multigranulation fuzzy rough set-based multi-attribute group decision making problems, Comput. Ind. Eng. 142 (2020) 106331, https://doi.org/10.1016/j.cie.2020.106331.
[10] K. Sun, S. Mou, J. Qiu, T. Wang, H. Gao, Adaptive fuzzy control for nontriangular structural stochastic switched nonlinear systems with full state constraints, IEEE Trans. Fuzzy Syst. 27 (8) (2019) 1587–1601, https://doi.org/10.1109/TFUZZ.2018.2883374.
[11] X. Yang, H. Chen, T. Li, P. Zhang, C. Luo, Student-t kernelized fuzzy rough set model with fuzzy divergence for feature selection, Inf. Sci. 610 (2022) 52–72, https://doi.org/10.1016/j.ins.2022.07.139.
[12] C.Y. Wang, L. Wan, New results on granular variable precision fuzzy rough sets based on fuzzy (co) implications, Fuzzy Sets Syst. 423 (2021) 149–169, https://doi.org/10.1016/j.fss.2020.08.011.
[13] Z. Yuan, H. Chen, T. Li, Z. Yu, B. Sang, C. Luo, Unsupervised attribute reduction for mixed data based on fuzzy rough sets, Inf. Sci. 572 (2021) 67–87, https://doi.org/10.1016/j.ins.2021.04.083.
[14] C. Wang, Y. Huang, W. Ding, Z. Cao, Attribute reduction with fuzzy rough self-information measures, Inf. Sci. 549 (2021) 68–86, https://doi.org/10.1016/j.ins.2020.11.021.
[15] Y. Lin, Y. Li, C. Wang, J. Chen, Attribute reduction for multi-label learning with fuzzy rough set, Knowl.-Based Syst. 152 (2018) 51–61.
[16] Z. Yuan, H. Chen, P. Xie, P. Zhang, J. Liu, T. Li, Attribute reduction methods in fuzzy rough set theory: an overview, comparative experiments, and new directions, Appl. Soft Comput. 107 (2021) 107353, https://doi.org/10.1016/j.asoc.2021.107353.
[17] J. He, L. Qu, Z. Wang, Y. Chen, D. Luo, C.-F. Wen, Attribute reduction in an incomplete categorical decision information system based on fuzzy rough sets, Artif. Intell. Rev. 55 (7) (2022) 5313–5348, https://doi.org/10.1007/s10462-021-10117-w.
[18] J. Dai, H. Hu, W.-Z. Wu, Y. Qian, D. Huang, Maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets, IEEE Trans. Fuzzy Syst. 26 (4) (2018) 2174–2187, https://doi.org/10.1109/TFUZZ.2017.2768044.
[19] X. Zhang, J. Wang, J. Zhan, J. Dai, Fuzzy measures and Choquet integrals based on fuzzy covering rough sets, IEEE Trans. Fuzzy Syst. 30 (7) (2022) 2360–2374, https://doi.org/10.1109/TFUZZ.2021.3081916.
[20] X. Zhang, J. Jiang, Measurement, modeling, reduction of decision-theoretic multigranulation fuzzy rough sets based on three-way decisions, Inf. Sci. 607 (2022) 1550–1582, https://doi.org/10.1016/j.ins.2022.05.122.
[21] W. Pedrycz, R. Al-Hmouz, A. Morfeq, A. Balamash, The design of free structure granular mappings: the use of the principle of justifiable granularity, IEEE Trans. Cybern. 43 (6) (2013) 2105–2113, https://doi.org/10.1109/TCYB.2013.2240384.
[22] Y. Yao, Three-way granular computing, rough sets, and formal concept analysis, Int. J. Approx. Reason. 116 (2020) 106–125, https://doi.org/10.1016/j.ijar.2019.11.002.
[23] J.T. Yao, A.V. Vasilakos, W. Pedrycz, Granular computing: perspectives and challenges, IEEE Trans. Cybern. 43 (6) (2013) 1977–1989, https://doi.org/10.1109/TSMCC.2012.2236648.
[24] G. Wang, J. Yang, J. Xu, Granular computing: from granularity optimization to multi-granularity joint problem solving, Granul. Comput. 2 (2017) 105–120, https://doi.org/10.1007/s41066-016-0032-3.
[25] D. Watson, H.F. Levin-Aspenson, M.A. Waszczuk, C.C. Conway, T. Dalgleish, M.N. Dretsch, N.R. Eaton, M.K. Forbes, K.T. Forbush, K.A. Hobbs, et al., Validity and utility of hierarchical taxonomy of psychopathology (hitop): iii. Emotional dysfunction superspectrum, World Psychiatry 21 (1) (2022) 26–54, https://doi.org/10.1002/wps.20943.
[26] Z. Xue, X. Tan, X. Yu, B. Liu, A. Yu, P. Zhang, Deep hierarchical vision transformer for hyperspectral and lidar data classification, IEEE Trans. Image Process. 31 (2022) 3095–3110, https://doi.org/10.1109/TIP.2022.3162964.
[27] S. Li, J. Yang, G. Wang, Q. Zhang, J. Hu, Granularity selection for hierarchical classification based on uncertainty measure, IEEE Trans. Fuzzy Syst. 30 (11) (2022) 4841–4855, https://doi.org/10.1109/TFUZZ.2022.3161747.
[28] Y. Wang, Q. Hu, H. Chen, Y. Qian, Uncertainty instructed multi-granularity decision for large-scale hierarchical classification, Inf. Sci. 586 (2022) 644–661, https://doi.org/10.1016/j.ins.2021.12.009.
[29] Z. Qiu, H. Zhao, A fuzzy rough set approach to hierarchical feature selection based on Hausdorff distance, Appl. Intell. 52 (2022) 11089–11102, https://doi.org/10.1007/s10489-021-03028-4.
[30] H. Zhao, P. Wang, Q. Hu, P. Zhu, Fuzzy rough set based feature selection for large-scale hierarchical classification, IEEE Trans. Fuzzy Syst. 27 (10) (2019) 1891–1903, https://doi.org/10.1109/TFUZZ.2019.2892349.
[31] S. Chang, R. Zhang, K. Ji, S. Huang, Z. Feng, A hierarchical classification head based convolutional gated deep neural network for automatic modulation classification, IEEE Trans. Wirel. Commun. 21 (10) (2022) 8713–8728, https://doi.org/10.1109/TWC.2022.3168884.
[32] F.E. Otero, A.A. Freitas, C.G. Johnson, A hierarchical multi-label classification ant colony algorithm for protein function prediction, Memetic Comp. 2 (2010) 165–181, https://doi.org/10.1007/s12293-010-0045-4.
[33] I. Dimitrovski, D. Kocev, S. Loskovska, S. Džeroski, Hierarchical annotation of medical images, Pattern Recognit. 44 (10–11) (2011) 2436–2449, https://doi.org/10.1016/j.patcog.2011.03.026.
[34] S. Chakrabarti, B. Dom, R. Agrawal, P. Raghavan, Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies, VLDB J. 7 (1998) 163–178, https://doi.org/10.1007/s007780050061.
[35] D. Dubois, H. Prade, Putting Rough Sets and Fuzzy Sets Together, Springer, Netherlands, Dordrecht, 1992, pp. 203–232.
[36] J. Yang, G. Wang, Q. Zhang, Knowledge distance measure in multigranulation spaces of fuzzy equivalence relations, Inf. Sci. 448–449 (2018) 18–35, https://doi.org/10.1016/j.ins.2018.03.026.
[37] B. Zhang, L. Zhang, Theory and Applications of Problem Solving, Elsevier Science Inc., New York, NY, USA, 1992.
[38] F.W. Lawvere, Metric spaces, generalized logic, and closed categories, Rend. Semin. Mat. Fis. Milano 43 (1973) 135–166, https://doi.org/10.1007/BF02924844.
[39] Uci repository. Available online http://archive.ics.uci.edu/ml/. (Accessed 10 October 2023), 2007.
[40] A. Konno, M. Inoue-Murayama, S. Yabuta, A. Tonoike, M. Nagasawa, K. Mogi, T. Kikusui, Effect of canine oxytocin receptor gene polymorphism on the successful training of drug detection dogs, J. Heredity 109 (5) (2018) 566–572, https://doi.org/10.1111/jeb.12230.
[41] S. Yavno, M.G. Fox, Morphological change and phenotypic plasticity in native and non-native pumpkinseed sunfish in response to sustained water velocities, J. Evol. Biol. 26 (11) (2013) 2383–2395, https://doi.org/10.1111/jeb.12230.
[42] P.A. Arnold, P. Cassey, C.R. White, Functional traits in red flour beetles: the dispersal phenotype is associated with leg length but not body size nor metabolic rate, Funct. Ecol. 31 (3) (2017) 653–661, https://doi.org/10.1111/1365-2435.12772.
[43] P. Tempski, I.S. Santos, F.B. Mayer, S.C. Enns, B. Perotta, H.B. Paro, S. Gannam, M. Peleias, V.L. Garcia, S. Baldassin, et al., Relationship among medical student resilience, educational environment and quality of life, PLoS ONE 10 (6) (2015) e0131535, https://doi.org/10.1371/journal.pone.0131535.
[44] M. Sebastiano, M. Eens, S. Messina, H. AbdElgawad, K. Pineau, G.T. Beemster, O. Chastel, D. Costantini, Resveratrol supplementation reduces oxidative stress and modulates the immune response in free-living animals during a viral infection, Funct. Ecol. 32 (11) (2018) 2509–2519, https://doi.org/10.1111/1365-2435.13195.
[45] Darius Danusevicius, 12 nuclear microsatellite loci scores for 543 adult trees of tilia cordata in Lithuainia, dryad, https://doi.org/10.5061/dryad.c866t1g64, 2022.
[46] L. Lacroix, S. Papis, C. Mardegan, F. Luterbacher, A. L'Huillier, C. Sahyoun, K. Keitel, N. Mastboim, L. Etshtein, L. Shani, et al., Host biomarkers and combinatorial scores for the detection of serious and invasive bacterial infection in pediatric patients with fever without source, PLoS ONE 18 (11) (2023) e0294032, https://doi.org/10.1371/journal.pone.0294032.

[47] J. Song, B. Huang, W. Qie, Allometry in late devonian podocopa ostracods (crustacea) and its implications for ostracod ontogeny, Pap. Palaeontol. 9 (1) (2023) e1480, https://doi.org/10.1002/spp2.1480.

[48] Y. Chen, T.A. Neubauer, L. Krystyn, S. Richoz, Allometry in a nisian (m iddle t riassic) segminiplanate conodonts and its implications for conodont taxonomy, Palaeontology 59 (5) (2016) 725–741, https://doi.org/10.1111/pala.12253.

[49] S. Bai, Y. Lin, Y. Lv, J. Chen, C. Wang, Kernelized fuzzy rough sets based online streaming feature selection for large-scale hierarchical classification, Appl. Intell. 51 (2021) 1602–1615, https://doi.org/10.1007/s10489-020-01863-5.